The determinants of regional freight transport: a spatial, semi-parametric approach

Tamás Krisztin*

Abstract In the context of modeling regional freight the four-stage model is a popular choice. The first stage of the model, freight generation and attraction, however, suffers from three shortcomings: first of all, it does not take spatial dependencies among regions into account, thus potentially yielding biased estimates. Second, there is no clear consensus in the literature as to the choice of explanatory variables. Secondly, sectoral employment and gross value added are used to explain freight generation, whereas some recent publications emphasize the importance of variables which measure the amount of logistical activity in a region. Third, there is a lack of consensus regarding the functional form of the explanatory variables. Multiple recent studies (Hesse and Rodrigue, 2004; Ranaiefar et al., 2013; Rodrigue, 2006; Tavasszy et al., 2012) emphasize nonlinear influences of selected variables. This paper addresses these shortcomings by using a spatial variant of the classic freight generation and attraction models combined with a penalized spline framework to model the explanatory variables in a semi-parametric fashion. Moreover, a Bayesian estimation approach is used, coupled with a penalized Normal inverse-Gamma prior structure, in order to introduce uncertainty regarding the choice and functional form of explanatory variables. The performance of the model is assessed on a real world example of freight generation and attraction of 258 European NUTS-2 level regions, covering 25 European countries.

Keywords: freight transportation, semi-parametric modeling, Bayesian econometrics, spatial autoregressive models

JEL Classification: C11, C14, C21, O18, R12, R40

October 12, 2016

1 Introduction

The four-stage model of regional transportation planning is a popular approach to assess the requirements for transportation and to adjust transportation policy. The model's first stage – freight generation and attraction – concerns the total volume of freight originating from or flowing to regions, respectively. This basic model has been a popular tool to explain and project the demand for transportation in a regional setting (see, e.g. Novak et al. 2008 or Ranaiefar et al.

---

*International Institute for Applied Systems Analysis, IIASA; E-mail address: krisztin@iiasa.ac.at

2013). In its original form – described, for example, in Ortúzar and Willumsen (2011) – freight generation and attraction models are simple linear regression models, where a region's sectoral, economic and demographic characteristics are regressed on the volume of freight generated and attracted, respectively, in the observation period.

This basic model of freight generation and attraction, however, suffers from a number of crucial shortcomings. First and foremost, it does not take the spatial nature of freight generation and attraction into account. Both in freight generation and attraction, spatial dependencies can arise, e.g. by neighboring regions influencing each others' volume of freight. These spatial dependencies can be the result of stronger economic ties between neighboring regions, e.g. if manufacturing plants are located in regions neighboring urban regions (De Grange et al., 2010). Recent modeling approaches in freight generation and attraction have found such spatial patterns. Hence Novak et al. (2008) and others use a spatial autoregressive model specification, introduced by Anselin (1988) to account for spatial lags of the variables. Their results indicate that spatial spillovers play a significant role in both freight generation and attraction.

A second problem associated with regional freight generation and attraction is the choice of explanatory variables in these models. The classic models include gross regional product and employment indicators for both freight generation and attraction. In the case of freight generation, sectoral gross value added and employment in manufacturing are considered to be relevant explanatory variables. For freight attraction, demand for final products can be proxied by the total population and intermediate demand is included as sectoral gross value added. These variables have been challenged by Hesse and Rodrigue (2004), Rodrigue (2006) and Tavasszy et al. (2012), among others. They argue for including variables that measure the regional intensity of logistic activities, in terms of the share of logistic companies or the average size of logistic and distribution companies. Novak et al. (2008), and Sánchez-Díaz et al. (2016) stress the importance of variables that characterize the geographical infrastructure, and suggest using the distance to ports or the length of highways in a region, while Chow et al. (2010), Boerkamps et al. (2000), Wisetjindawat and Sano (2003) and Wagner (2010) draw attention to land-use indicators as proxy for logistical activity, arguing that a higher share of developed surfaces in a region indicates a higher volume of generated and attracted freight. Nevertheless, it seems unclear, which explanatory variables to include in models of freight generation and attraction.

Including a full selection of all proposed explanatory variables would substantially reduce the degrees of freedom in a regional economic setting and potentially cause problems with multicollinearity of the explanatory variables. Multiple studies (Ranaiefar et al., 2013; Tavasszy et al., 2012) in freight generation address this issue by manually selecting the subset of variables which yield the best chosen measure of fit (for example using R^2 or the Akaike Information Criterion as measures). Such an approach might not be optimal, as – even with a moderate number of co-variates – testing all possible combinations might not be feasible. Only testing a subset of the models might lead to biased choices of variables (Fahrmeir et al., 2004).

Such variable selection issues may be addressed through using Bayesian techniques (LeSage and Parent, 2007; Piribauer and Fischer, 2015). Instead of having to select a single set of explanatory variables, Bayesian variable selection approaches advocate using a model with a wide selection of different variables, and through an adequate selection of priors, using the posterior importance of variables for inference. In this fashion a large number of potential explanatory variables can be included and parameter inference can be made unconditional on model specifications. There is a wide array of literature on variable selection in non-spatial frameworks (e.g. see Koop 2003, George and McCulloch 1993 and Kuo and Mallick 1998), and various approaches to incorporate variable selection into a spatial framework (see, for example, Lesage and Fischer, 2009).

A third shortcoming of the freight generation and attraction models lies in the uncertainty over which functional form to use for the explanatory variables. Sánchez-Díaz et al. (2016) suggest multiple nonlinear transformations (for example logarithmic, quadratic or exponential transformations, as in Novak et al. 2008) of the explanatory variables, until a sufficient value of the chosen measure of fit is achieved. As noted by Tavasszy et al. (2012) and Rodrigue (2006) such an exploratory approach shows some drawbacks: first, it is difficult to specify which functional form would be the most appropriate for each variable. Second, testing for a wide number of transformations can be computationally burdensome and including polynomials of high order can lead to numerical instability.

Such nonlinearities in the parameters, however, seem to play a central role in freight generation and attraction. Even in the case of freight generation and attraction models that account for spatial dependence the transformation of variables can improve the model fit (Novak et al., 2008). Chow et al. (2010), and De Jong et al. (2013) among others address this issue by means of a semi-parametric approach. Ranaiefar et al. (2013) suggest using a structural equation modelling approach to capture these effects, their model, however suffers from a number of key weaknesses. First, while they take the spatial nature of freight generation into account, they only include spatial lags of the explanatory variables, not of the dependent variable. Second, their structural equation model does not explicitly differentiate between spatial correlation and nonlinearities in the data, instead it subsumes this into joint correlation estimates, which capture any additional nonlinearity. Moreover, structural equation modelling usually requires large data samples to perform adequately, which makes it ill suited for the modelling of regional freight generation.

Semi-parametric modeling in the context of spatial autoregressive models was recently addressed by Basile (2008), Del Bo and Florio (2012), Fotopoulos (2012) and Basile et al. (2014). The advantage of the semi-parametric approach, relying on the parametric expansion of co-variates through spline functions, is that it can be easily used in conjunction with classical spatial econometric estimation approaches. The main drawback of this approach is that in a frequentist context, it is often difficult to estimate, due to the increased number of effective parameters[1]. While frequentist solutions such as penalized least-squares exist (Eilers and Marx, 1996), these are typically outperformed by Bayesian estimation methods relying on some form of shrinkage priors (Lang and Brezger, 2004). Moreover, Bayesian estimation methods can potentially be combined with the aforementioned Bayesian variable selection approaches. This was recently proposed (albeit in a non-spatial context) in Brezger and Lang (2006); Fahrmeir et al. (2004); Lang and Brezger (2004) by using random-walk priors for the splines, in conjunction with spike-and-slab priors for the selection of explanatory variables.

This paper aims to address three central issues associated with freight generation and attraction modeling. *First*, it is increasingly recognized that freight generation (and attraction) is a spatial process, characterized by spatial dependencies. Neglecting these dependencies would lead to biased and incorrect estimates. To address this issue, our approach takes spatial lags of the dependent and the explanatory variables into account. *Second*, we address the issue of uncertainty arising from explanatory variable selection. For this purpose we use a Bayesian variable selection framework, using spike-and-slab priors, which can consider models including a potentially large number of co-variates and provides posterior evidence on the relative importance of explanatory variables. *Third*, it is the linearity in parameters in conventional linear regression models that greatly simplifies the analysis of this class of models. But it also leads to some significant limitations. Since previous literature suggests that such nonlinearities could play an important role in the modeling of freight generation, we relax the assumption of linearity through a novel Bayesian semi-parametric modeling approach. This enables us to test for the presence or absence of nonlinearity in the impact of the explanatory variables.

In summary, this paper contributes in multiple ways to the literature: it presents a novel spatial, semi-parametric modeling approach in conjuction with variable selection. To our knowledge, no other modeling approach allows for simultaneous inference on the relative importance of co-variates, their functional form, as well as the strength of spatial dependencies of the response variable. Moreover, in the context of freight generation and attraction modeling, this paper addresses the questions, whether there is significant spatial dependence in the volume of freight generated by (or attracted to) European regions, what the central explanatory variables of freight generation (or attraction) are, and finally whether these variables have a nonlinear impact on freight generation (or attraction).

The rest of the paper is organized as follows. Section 2 presents the basic freight generation and attraction model, together with its spatial counterpart. Section 3 expands the spatial version of the model by introducing nonlinearity in the parameters through basic splines and discussing the penalized spline structure. In section 4 we continue the prior set-up, and present the full

---

[1]Consider, that a parametric expansion relies on splitting each co-variate into a set of basis functions, which effectively multiplies the number of co-variates by the number of basis functions used for the expansion. Even a comparatively low number of basis functions (e.g. five to ten), would result in five to ten times as many parameters in the semi-parametric model.

conditional posteriors and the sampling procedure employed in the Bayesian estimation of the model.

Section 5 serves to test the proposed model approach using an empirical example of freight generation and attraction with 258 European NUTS-2 level regions, that cover the EU-27 countries, excluding Cyprus and Malta. We model the average freight generation and attraction in the period 2010-2014, based on a selection of 20 covariates and their spatially lagged counterparts. Out of these variables we model 14 (and their spatially lagged counterparts) in a semi-parametric fashion. The results indicate that spatial spillovers play a significant role, and that in the case of sectoral manufacturing variables a nonlinear parametrization should be used. Section 6 concludes.

2 The spatial freight generation and attraction models

Let y_i denote the volume of freight generated by (or attracted to) region i ($i = 1, ..., N$), where N is the number of regions in the study area. The model postulates that y_i can be explained by a $1 \times K$ vector $\boldsymbol{x}_i$, containing observations on K socio-economic characteristics, with an K by 1 vector of associated coefficients $\boldsymbol{\theta}$, an intercept θ_0, and an error term ε_i, assumed to be normally and identically distributed for all i, with zero mean and σ^2 variance. Thus we can write:

$$y_i = \theta_0 + \boldsymbol{x}_i \boldsymbol{\theta} + \varepsilon_i \tag{1}$$
$$\varepsilon_i = \mathcal{N}(0, \sigma^2).$$

The linear regression model in Eq. (1) is referred to as the basic freight generation (or attraction) model (De Grange et al., 2013).

In freight generation and attraction models spatial dependence may arise in two ways, first, by neighboring regions influencing the volume of freight generated or attracted by a given region, and second by spatially dependent explanatory variables - such as land-use or population characteristics (Anselin and Bera, 1998; Arbia, 2006). To account for these spatial dependencies we use a spatial Durbin model (SDM) specification (LeSage and Pace, 2009). This can be written as:

$$y_i = \rho \sum_{q=1}^{N} w_{i,q} y_q + \theta_0 + \boldsymbol{x}_i \boldsymbol{\theta} + \sum_{q=1}^{N} w_{i,q} \boldsymbol{x}_q \boldsymbol{\lambda} + \varepsilon_i. \tag{2}$$

$w_{i,q}$ (with $q = 1, ..., N$) is a typical element of the N by N exogenously given non-negative spatial weight matrix $\mathbf{W}$. If observations i and q are considered to be neighbors, then $w_{i,q} > 0$, otherwise $w_{i,q} = 0$, with $w_{i,q|i=q} = 0$ (meaning that no observation is a neighbor to itself). $\mathbf{W}$ is assumed to be row-stochastic, so that $\sum_q w_{i,q} = 1 \ \forall q = 1, \ldots, N$ (Fischer and Wang, 2011).

The term $w_{i,q} y_q$ describes an endogenous interaction effect among the dependent variable, the term $w_{i,q} \boldsymbol{x}_q$ exogenous interaction effect among the independent variables. $\rho \in [-1, 1]$ is called the spatial autoregressive coefficient, while $\boldsymbol{\lambda}$ – just as $\boldsymbol{\theta}$ – represents a $K \times 1$ vector of fixed, but interlinked parameter vector. Note that Eq. (2) subsumes Eq. (1) as a special case, if $\rho = 0$ and $\lambda_1, \ldots, \lambda_K = 0$.

To simplify our notation, we rewrite Eq. (2) in matrix notation. Let $\boldsymbol{y} = [y_1, ..., y_N]'$ and let the N by $2K$ matrix $\mathbf{V} = [\mathbf{X}, \mathbf{W}\mathbf{X}]$, where $\mathbf{X} = [\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N]'$ is the N by K matrix of explanatory variables. In this way we can re-write Eq. (2) as:

$$\boldsymbol{y} = \rho \mathbf{W} \boldsymbol{y} + \boldsymbol{\iota}_N \theta_0 + \mathbf{V} \boldsymbol{\phi} + \boldsymbol{\varepsilon} \tag{3}$$
$$\boldsymbol{\varepsilon} = \mathcal{N}(0, \sigma^2 \mathbf{I}_N)$$

where $\boldsymbol{\phi} = [\boldsymbol{\theta}', \boldsymbol{\lambda}']'$, $\boldsymbol{\iota}_N$ denotes an $N \times 1$ vector of ones, and $\mathbf{I}_N$ is an $N \times N$ identity matrix.

One important feature of models accounting for spatial dependence is that the interpretation of the partial derivatives becomes richer but more complicated. In classical freight generation and attraction models, such as in Eq. (1), the partial derivatives of y_i with respect to x_{ik} have the simple form where $\frac{\partial y_i}{\partial x_{ik}} = \theta_k$ and $\frac{\partial y_i}{\partial x_{jk}} = 0$ for all $i \neq j$ and all $k = (1, \ldots, K)$ variables. However in the spatial dependence case with $\rho \neq 0$ and $\lambda_k \neq 0$, the interpretation of the partial derivatives becomes richer but more complicated. We follow (LeSage and Pace, 2009) and use summary impact measures to asses the average direct, indirect and total impact of changes in the k-th $\boldsymbol{x}$ model variable.

4

3 Semi-parametric modeling of freight generation and attraction

While the model in Eq. (3) can capture nonlinearities associated by spatial patterns, as pointed out by Rodrigue (2006) freight generation models may contain further nonlinearities in the explanatory variables as well, which are non-spatial in nature. We address this issue by following a strain of literature from regional economic growth models, which combines spatial models and semi-parametric modeling. This approach, pioneered by Basile (2008); Basile et al. (2014); Del Bo and Florio (2012); Fotopoulos (2012), is characterized by the use of a univariate smoothing function that allows a more flexible modeling of the explanatory variables and thus capture potential nonlinearities in the parameters. These papers use a form of parameter expansions called basic splines[2] (B-splines), based on locally defined piece-wise polynomials, to model explanatory variables in a flexible way (Ruppert et al., 2003).

Semi-parametric modeling via B-Splines

To capture potential nonlinear influences of the explanatory variables on the dependent variable in Eq. (3) we assume a continuous function $f_m(.)$ for the set of explanatory variables:

$$\boldsymbol{y} = \rho \mathbf{W} \boldsymbol{y} + \boldsymbol{\iota}_n \theta_0 + \boldsymbol{f}(\mathbf{V}) + \boldsymbol{\varepsilon} \tag{4}$$

$$\boldsymbol{f}(\mathbf{V}) = \sum_{m=1}^{2K} f_m(\boldsymbol{v}_m)$$

where $\boldsymbol{v}_m$ (with $m = 1, \ldots, 2K$) denotes the m-th column of $\mathbf{V}$ and the exact functional form of each $f_m(\cdot)$ is unknown. $f_m(\cdot)$ is assumed to be a smooth function with continuous first order derivatives. Continuous derivatives are essential for calculating the direct and indirect spatial summary impact measures.

Following Lang and Brezger (2004), we model the unknown functions $f_m(\cdot)$ through B-Splines, introduced by Eilers and Marx (1996). The core assumption underlying this approach is that we can approximate $f_m(\cdot)$ through a set of piecewise polynomials, each of D_m-th degree, and defined over P_m knots, denoted as $\boldsymbol{\chi}_m$. The L_m elements of $\boldsymbol{\chi}_m$ are equally spaced over the range spanned by $min(\boldsymbol{v}_m)$ and $max(\boldsymbol{v}_m)$ [where $min(\boldsymbol{v}_m)$ denotes the smallest and $max(\boldsymbol{v}_m)$ the largest element of $\boldsymbol{v}_m$, respectively]. Each knot $\chi_{m,p_m} \in \boldsymbol{\chi}_m$ (with $p_m = 1, \ldots, P_m$) is a support point for the p_m-th piecewise polynomial. Each piecewise polynomial is defined over the range of exactly D_m knots and are zero otherwise. If we let the lp_m-th column of the $N \times P_m$ design matrix $\bar{\mathbf{Z}}_m$ correspond to the splines representation of the p_m-th polynomial (as defined in the seminal work by DeBoor 1978; for details see Appendix A, we can rewrite $f_m(\cdot)$, so that:

$$f_m(\boldsymbol{v}_m) = \bar{\mathbf{Z}}_m \bar{\boldsymbol{\beta}}_m. \tag{5}$$

The unknown function $f_m(\boldsymbol{v}_m)$ models m-th explanatory variable in a semi-parametric fashion, through a so-called design matrix $\bar{\mathbf{Z}}_m$ and $\bar{\boldsymbol{\beta}}_m$, a $P_m \times 1$ vector of corresponding coefficients. The bars above the terms $\bar{\mathbf{Z}}_m$ and $\bar{\boldsymbol{\beta}}_m$ signify that the terms are associated with parameter expanded basis functions. The number of columns of the design matrix $P_m = L_m + 2D_m$ corresponds exactly to the number of knots and twice the degree of the spline. The D_m extra columns, termed as so-called support knots in the spline literature, define the spline outside of $\boldsymbol{v}_m^{min}$ and $\boldsymbol{v}_m^{max}$. For details, see Appendix A.

The derivatives of a B-spline, represented through P_m piecewise polynomials of degree D_m over P_m support knots, can in turn be expressed by piecewise polynomials of degree $D_m - 1$ over the same support knots. Similar to the representation of the B-spline above, we can express its first order derivatives in terms of an $N \times P_m$ design matrix $\bar{\mathbf{Z}}_m^\partial$ and the vector of coefficient $\bar{\boldsymbol{\beta}}_m$, where ∂ signifies that the matrix $\bar{\mathbf{Z}}_m^\partial$ represents first-order derivatives. For details on the construction of the matrix $\bar{\mathbf{Z}}_m^\partial$ of partial derivatives, see DeBoor (1978).

---

[2]Basic splines are a class of semi-parametric basis function, which can be used to model nonlinearities in the parameter. This is achieved by using a large set of overlapping piecewise polynomials (the so-called bases) to model each explanatory variables (DeBoor, 1978).

Based on the fully specified representation of $f_m(\cdot)$ in Eq. (5), we can rewrite Eq. (4) as:

$$\boldsymbol{y} = \rho\mathbf{W}\boldsymbol{y} + \boldsymbol{\eta}_0 + \sum_{m=1}^{2K} \bar{\mathbf{Z}}_m \bar{\boldsymbol{\beta}}_m + \boldsymbol{\varepsilon} \tag{6}$$

where $\boldsymbol{\eta}_0 = \mathbf{X}_0 \boldsymbol{\beta}_0$. The $N \times K_0$ matrix $\mathbf{X}_0$ contains the intercept and other variables which we do not want to model in a semi-parametric fashion (for example fixed or random effects). $\boldsymbol{\beta}_0$ is the $K_0 \times 1$ vector of corresponding coefficients. [3]

It is easy to see from Eq. (6) that the B-spline model is similar to the SAR model discussed in the previous section, but with a large number of additional coefficients. The calculation of summary statistics for direct, indirect and total partial derivatives presents a slight departure from the SAR model, since the first-order derivatives of B-splines are not equal to the coefficients. Let the modified total derivative matrix, which includes the correct derivatives for B-splines, be denoted as $\mathbf{S}_k(\mathbf{W})$. Then, we can write:

$$\frac{\partial y_i}{\partial x_{jk}} = \mathbf{S}_k(\mathbf{W})_{ij} \text{ and } \frac{\partial y_i}{\partial x_{ik}} = \mathbf{S}_k(\mathbf{W})_{ii}$$
$$\mathbf{S}_k(\mathbf{W}) = (\mathbf{I}_n - \rho\mathbf{W})^{-1} \left[diag\left(\frac{\partial f_k(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k}\right) + diag\left(\frac{\partial f_k(\mathbf{W}\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k}\right) \right] \tag{7}$$

where the $diag(\cdot)$ operator transforms a vector into a matrix, with the vectors' values along the main diagonal. Based on $\mathbf{S}_k(\mathbf{W})$, we can compute the summary direct, indirect and total effects in the same fashion as in LeSage and Pace (2009).

The B-spline model can be estimated in a similar fashion as the SAR model. It is, however, very likely that the model in Eq. (6) is over-specified, since the total number of coefficients $(\sum_{m=1}^M P_m + K_0 + 1)$ is quite likely to come close or even exceed N. This is obviously dependent on the total number of support knots $\sum_m P_m$. If we choose a large number (for example ten or more) of equally spaced knots, we are rather flexible in the choice of functional form for $\boldsymbol{f}(\cdot)$. If we only use relatively few knots, we might be too inflexible and our particular parametrization might not adequately capture prevalent nonlinearities. While the B-spline representation in Eq. (6) clearly allows for additional flexibility in modeling the impacts of the explanatory variables, this comes at the cost of the so-called curse of dimensionality; the more flexible our modeling approach, the more observations would be needed. In most regional studies, observations are limited to cross-sectional and panel data contexts, usually comprising not more than 1,000 observations. This, of course, severely restricts the number of covariates that could be modeled with simple B-splines alone.

Penalized splines

The ability of splines to correctly approximate the continuous nonlinear function $f_m(\cdot)$ is dependent on the number of spline knots L_m. Furthermore, selecting a large number of spline knots raises the issues of over-parametrization and over-fitting, while selecting too few could mean that we do not approximate $f_m(\cdot)$ to a sufficient degree using B-splines. Multiple approaches have been proposed to deal with this problem. As suggested by Koop and Poirier (2004), one could vary the number of design knots in order to minimize certain criteria, for example the AIC or the Bayes' factor. This approach, however, usually show rather slow convergence, if the number of explanatory variables is increased. To circumvent this problem, we allow for a relatively large number of uniformly placed knots and follow Eilers and Marx (1996) to utilize a penalized estimation procedure, termed P-splines.

This approach applies the concepts of Bayesian shrinkage to the B-spline model and assumes a random walk structure for the parameter vector $\bar{\boldsymbol{\beta}}_m$ (for $1, \dots, M$), where the conditional expectation of $\bar{\boldsymbol{\beta}}_{p_m,m}$ is dependent on $\bar{\boldsymbol{\beta}}_{p_m-1,m}$ (Eilers and Marx, 1996). More formally, and assuming a random walk structure of order one:

$$p(\bar{\boldsymbol{\beta}}_{p_m,m}|\bar{\boldsymbol{\beta}}_{p_m-1,m}, ..., \bar{\boldsymbol{\beta}}_{1,m}) \sim \mathcal{N}(\bar{\boldsymbol{\beta}}_{p_m-1,m}, \tau_m^2) \quad p_m = 1, \dots, P_m \tag{8}$$

---

[3]Note that in Eq. (6) the mean levels of the smooth functions $f_m(\cdot)$ are not identifiable. To ensure identifiability all basis functions $f_m(\cdot)$, and thus all design matrices $\bar{\mathbf{Z}}_m$ are constrained to have zero mean.

where the coefficient τ_m^2 is a measure of the m-th random walk's variance. On the one hand, if τ_m^2 is large, the B-spline coefficients associated with each of the p_m piecewise polynomials in $\bar{\mathbf{Z}}_m$ are allowed to differ to a large degree, which leads to larger variances in the functional form of the resulting splines. If, on the other hand, τ_m^2 is relatively small, the coefficients in $\bar{\boldsymbol{\beta}}_m$ will be very closely associated with each other, which leads to a nearly linear functional form for $f_m(\cdot)$.

This random walk structure for $p(\bar{\boldsymbol{\beta}}_m | \tau_m^2)$ can be incorporated by reparametrizing the design matrices $\bar{\mathbf{Z}}_m$ in Eq. (6), so that they include the random walk specification (Scheipl et al., 2012). We use the reparametrization based on spectral decomposition, as suggested in Fahrmeir et al. (2004) and earlier semiparametric approaches, such as in Lang and Brezger (2004), and Ruppert et al. (2003). This procedure relies on decomposing $f(\cdot)$, into the separate function $f_0(\cdot)$ and $f_{pen}(\cdot)$. The function $f_0(\cdot)$ spans the null-space of $f(\cdot)$ and can be interpreted as it's linear part component. $f_{pen}(\cdot)$ represents the penalized part of $f(\cdot)$, and can be interpreted as it's nonlinear components. Based on this reparametrization approach, and coupled with Bayesian variable selection priors, we can infer whether the linear or the nonlinear, or both parts of the splines have a higher posterior probability of being included in our models.

Based on this reasoning, let us assume that reparametrization has been applied to all $2K$ semiparametric model terms, thus yielding $4K$ model terms. We might, moreover, wish to include K_{lin} explanatory variables as model terms, which should not be modeled in a semi-parametric fashion, but we still want to make inference over their inclusion probability. Together, this adds up to $Q = 4K + 2K_{lin}$ separate model terms, where we denote the j-th model term (whether it is a penalized or unpenalized matrix) by the $n \times P_j$ design matrix $\mathbf{Z}_j$, with $j = 1, \ldots, Q$, with corresponding $P_j \times 1$ parameter vector $\boldsymbol{\beta}_j$. Thus we can write

$$\boldsymbol{y} = \rho \mathbf{W} \boldsymbol{y} + \boldsymbol{\eta}_0 + \sum_{j=1}^{Q} \mathbf{Z}_j \boldsymbol{\beta}_j + \boldsymbol{\varepsilon} \tag{9}$$

where, due to the reparametrization, $p(\boldsymbol{\beta}_j | \tau_j^2) \sim \mathcal{N}(0, \tau_j^2 \mathbf{I}_N)$.

To sum up, the reparametrization serves three main purposes: first, it naturally incorporates the above random walk prior into the model formulation itself; second, it enables us to separate each smooth polynomial term into an unpenalized and a penalized part. This in turn makes it possible to evaluate whether a penalized term should be included in the model at all, or whether the null-space of the explanatory variable (i.e. in the case of a first order random walk penalization structure its linear form) is sufficient. Thus we can obtain inference over the linearity of individual explanatory variables. Third, the reparametrization makes prior elicitation easier, since the reparametrized term has a proper Gaussian distribution, whereas the random-walk parametrization is improper (Scheipl et al., 2012).

4 Bayesian estimation

We follow a Bayesian approach in estimating Eq. (9). This is motivated by the usefulness of penalized splines in circumventing problems of dimensionality through a random-walk prior as in Eq. (8) and its re-parameterized counterpart. Moreover, a Bayesian approach allows us to make use of so-called variable selection priors. Such a prior structure allows for inference over the posterior inclusion probability of our explanatory variables. Thus, we can easily address the issue of uncertainty, whether, for example, supply-chain variables influence freight generation and attraction. We incorporate this into our model by using a penalized Normal inverse-Gamma prior set-up, in the spirit of Scheipl et al. (2012).

Prior set-up

To conduct Bayesian inference we have to specify prior distributions for all parameters in the model. Our specific prior set-up follows the penalized spline approach of Scheipl et al. (2012) in implementing the penalized Normal inverse-Gamma prior. This prior set-up is characterized by using a multiplicative parametrization for $\boldsymbol{\beta}_j$, motivated by the need to choose specific explanatory variables and to decide whether a co-variate should be modeled in a linear or semi-parametric

fashion. Specifically, we use a so-called spike-and-slab prior, combined with the Normal-Gamma prior on τ_j^2. We achieve this by multiplicatively expanding $\boldsymbol{\beta}_j = \alpha_j \boldsymbol{\zeta}_j$, where the scalar α_j expresses the importance of the j-th co-variate, while the $P_j \times 1$ vector $\boldsymbol{\zeta}_j$ serves to distribute α_j across the parameter block $\boldsymbol{\beta}_j$. The prior and hyperparameter structure for α_j is as follows:

$$p(\alpha_j) \sim \mathcal{N}\left(0, \tau_j^2\right) \;\; \text{with} \;\; \tau_j^2 = \gamma_j \nu_j^2 \tag{10}$$

$$p(\nu_j^2) \sim \mathbf{\Gamma^{-1}}\left(\underline{a}_\nu, \underline{b}_\nu\right) \tag{11}$$

$$p(\gamma_j) \sim \omega \delta_1\left(\gamma_j\right) + (1-\omega)\,\delta_{\kappa_0}\left(\gamma_j\right) \tag{12}$$

$$p(\omega) \sim \mathcal{B}\left(\underline{a}_\omega, \underline{b}_\omega\right) \tag{13}$$

The prior on the scalar α_j in Eq. (10) is motivated by the assumption that α_j follows a-priori a Gaussian distribution, with zero mean and variance τ_j^2. The variance τ_j^2 is split into two components: γ_j and ν_j^2. γ_j is a shrinkage indicator variable and ν_j^2 is the prior variance. We assume an additional level of prior hierarchy and set as the prior for ν_j^2 an inverse Gamma distribution, with the parameters $\underline{a}_\nu$ and $\underline{b}_\nu$, and $\underline{a}_\nu \ll \underline{b}_\nu$ [see Eq. (11)].

γ_j is assumed to take the value of one with probability ω and some very small value κ_0 with the probability $1-\omega$. $\delta_{\kappa_0}(\gamma_j)$ is zero for any $\gamma_j \neq \kappa_0$, and is one if $\gamma_j = \kappa_0$. The prior in Eq. (12) for γ_j completes the spike and slab prior on α_j. The implied prior on the variance $\tau_j^2 = \gamma_j \nu_j^2$ has one part of its prior mass concentrated on very small values close to zero – called the spike, with $\gamma_j = \kappa_0$ — and the other part of the prior mass being more diffuse and centred on larger values – termed as the slab, with $\gamma_j = 1$. An explanatory variables' posterior coefficient γ_j, that is strongly sampled from the spike prior mass, will be aggressively shrunk towards zero (if κ_0 is set to a sufficiently small value). In this fashion, the posterior probability of $\gamma_j = \kappa_0$ can be interpreted as the probability of exclusion of $f_j(\cdot)$ from the model. The Beta distributed hyperprior on ω in Eq. (13), with parameters $\underline{a}_\omega$ and $\underline{b}_\omega$ sets the overall probability of explanatory variables being ex- or included in the model.

For the spline coefficient vector $\boldsymbol{\zeta}_j$, we set a Gaussian prior distribution, shown in Eq. (14), with mean $g_{l_j,j}$ and variance of one:

$$p(\boldsymbol{\zeta}_j|\boldsymbol{g}_j) \sim \mathcal{N}\left(\boldsymbol{g}_j, \mathbf{I}_Q\right) \tag{14}$$

$$p(g_{l_j,j}) \sim \frac{1}{2}\delta_1\left(g_{l_j,j}\right) + \frac{1}{2}\delta_{-1}\left(g_{l_j,j}\right). \tag{15}$$

The mean $g_{l_j,j}$ of the Gaussian prior distribution for $\boldsymbol{\zeta}_j$ is with equal probability $+1$ or -1, as in Eq. (15). This effectively assigns as a prior for $\boldsymbol{\zeta}_j$ a mixture of two independently and identically distributed Gaussian mixture distributions with prior mean of ± 1. While the mean of $p(\boldsymbol{\zeta}_j)$ is still equal to zero, the bivariate prior assigns most of the prior mass to either positive or negative unity, thereby enabling us to interpret α_j as the relative importance of the j-th explanatory variable (Scheipl et al., 2012).

The prior structure given by Eq. (10) – Eq. (15) may be termed the penalized Normal mixture inverse-Gamma prior. Note, that for the spike-and-slab part of the prior structure in Eq. (12) and Eq. (13), we used the spike and slab approach from Ishwaran and Rao (2005). An alternative spike and slab prior, which has been suggested for similar models, is the stochastic search variable selection (SSVS) prior by George and McCulloch (1997), as well as the prior specification by Kuo and Mallick (1998), which was recently suggested for use in spatial autoregressive models by Piribauer and Cuaresma (2016).

The key difference of the prior structure of Ishwaran and Rao (2005) compared to the SSVS or the prior structure of Kuo and Mallick (1998) is the use of a continuous Beta distributed prior for the hyperparameter ω in Eq. (13), compared to the bimodal binomial prior that the SSVS prior uses. This implies that the prior is somewhat easier to specify than in the SSVS case. This is due to the fact that the classic implementation of the SSVS prior requires running a Maximum Likelihood estimate of the model, in order to estimate the variance of the prior distributions. This proves a difficult task in conjunction with splines, since the number of effective co-variates makes Maximum Likelihood estimation often infeasible. The prior set-up by Ishwaran and Rao (2005), is independent of this estimation procedure. While the prior structure suggested by Kuo and Mallick (1998) does not necessitate a previous estimation of the prior parameters, but it still requires

specifying a seperate ω for each co-variate, as opposed to the prior structure of Ishwaran and Rao (2005). Additionally, Ishwaran and Rao (2005) provide some desirable analytical properties of a Beta prior for ω, especially if the hyperparameters $\underline{a}_\omega$ and $\underline{b}_\omega$ are set, so that the distributions is close to uniform.

For the prior on $\boldsymbol{\beta}_0$, we follow the canonical Bayesian approach with a Gaussian prior [see Eq. (16)], with $\underline{\boldsymbol{\mu}}_{\beta_0}$ vector of prior means and $\underline{\boldsymbol{\Sigma}}_{\beta_0}$ as prior variance:

$$p(\boldsymbol{\beta}_0) \sim \mathcal{N}\left(\underline{\boldsymbol{\mu}}_{\beta_0}, \underline{\boldsymbol{\Sigma}}_{\beta_0}\right) \tag{16}$$

$$p(\sigma^2) \sim \boldsymbol{\Gamma}^{-1}\left(\underline{a}_{\sigma^2}, \underline{b}_{\sigma^2}\right). \tag{17}$$

The prior for the model variance σ^2 is an inverse-Gamma distribution with parameters $\underline{a}_{\sigma^2}$ and $\underline{b}_{\sigma^2}$. Finally, for the prior of ρ, we follow LeSage and Pace (2009):

$$p(\rho) \sim \mathcal{U}\left(-1, 1\right) \tag{18}$$

and assign a uniform prior in the range of ± 1. This concludes our prior set-up.

Conditional posteriors

The prior structure described in the previous section combined with the full model likelihood gives rise to a set of conditional posteriors from well-known distributions, all available in closed form, which facilitates a Markov Chain Monte Carlo (MCMC) algorithm.

For notational convenience let $\mathbf{Z} = [\mathbf{Z}_1, \ldots, \mathbf{Z}_Q]$ denote the full set of design matrices. The full conditional posterior for the multiplicative parameter vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_Q)'$ depends on design matrix $\mathbf{Z}_\alpha = \mathbf{Z}\, blockdiag\left(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_Q\right)$, conditional on $\boldsymbol{\zeta} = \left(\boldsymbol{\zeta}_1', \ldots, \boldsymbol{\zeta}_Q'\right)$. Analogously, $\mathbf{Z}_\zeta$ depends on the design matrix $\mathbf{Z}_\zeta = \mathbf{Z}\, diag\left(blockdiag\left[\boldsymbol{\iota}_{L11}, \ldots, \boldsymbol{\iota}_{LQQ}\right]\boldsymbol{\alpha}\right)$. Furthermore, let $\mathbf{A} = (\mathbf{I}_N - \rho\mathbf{W})^{-1}$.

Under the prior assumption given in Eqs. (10) - (13) the conditional posterior for $\boldsymbol{\alpha}$ is given by

$$p(\boldsymbol{\alpha}|\cdot) \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\right)$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \left(\frac{1}{\sigma^2}\mathbf{Z}_\alpha'\mathbf{Z}_\alpha + \mathrm{diag}\left(\boldsymbol{\gamma}\boldsymbol{\nu}^2\right)^{-1}\right)^{-1} \tag{19}$$

$$\boldsymbol{\mu}_{\boldsymbol{\alpha}} = \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\left(\frac{1}{\sigma^2}\mathbf{Z}_\alpha'\mathbf{A}\boldsymbol{y}_{spl}\right).$$

Here, the $N \times 1$ vector $\boldsymbol{y}_{spl}$ in Eq. (19) corresponds to the response $\boldsymbol{y}$ without the linear trend $\boldsymbol{\eta}_0$. The full conditional posterior densities for γ_j and ν_j^2 given as:

$$\frac{p(\gamma_j = 1|\cdot)}{p(\gamma_j = \kappa_0|\cdot)} = \frac{\omega}{1-\omega}\kappa_0^{1/2}\exp\left(\frac{1-\kappa_0}{2\kappa_0}\frac{\boldsymbol{\alpha}_j^2}{\nu_j^2}\right) \tag{20}$$

$$p(\nu_j^2|\cdot) \sim \boldsymbol{\Gamma}^{-1}\left(\underline{a}_\nu + \frac{1}{2}, \underline{b}_\nu + \frac{\boldsymbol{\alpha}_j^2}{2\gamma_j}\right). \tag{21}$$

The posterior for the hyperparamater ω is given as:

$$p(\omega|\cdot) \sim \mathcal{B}\left(\underline{a}_\omega + \sum_j^Q \delta_1(\gamma_j), \underline{b}_\omega + \sum_j^Q \delta_{\kappa_0}(\gamma_j)\right) \tag{22}$$

where $\delta_1(\cdot)$ and $\delta_{\kappa_0}(\cdot)$ correspond to the Dirac-delta function, with parameter one and κ_0, respectively. Given the priors in Eqs. (14) - (15), the conditional posterior of $\boldsymbol{\zeta}$ is Gaussian and given as:

$$p(\boldsymbol{\zeta}|\cdot) \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\zeta}}, \boldsymbol{\Sigma}_{\boldsymbol{\zeta}}\right)$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\zeta}} = \left(\frac{1}{\sigma^2}\mathbf{Z}_\zeta'\mathbf{Z}_\zeta + \mathbf{I}_N\right)^{-1}$$

$$\boldsymbol{\mu}_{\boldsymbol{\zeta}} = \boldsymbol{\Sigma}_{\boldsymbol{\zeta}}\left(\frac{1}{\sigma^2}\mathbf{Z}_\zeta'\mathbf{A}\boldsymbol{y}_{lin} + \boldsymbol{g}\right)$$

where $\boldsymbol{g} = [\boldsymbol{g}'_1, \ldots, \boldsymbol{g}_j, \ldots, \boldsymbol{g}'_Q]'$ and $\boldsymbol{g}_j = [g_1, \ldots, g_{l_j}, \ldots, g_{L_j}]$. Furthermore, let $\boldsymbol{y}_{lin} = \boldsymbol{y} - \sum_{j=1}^{Q} \mathbf{Z}_j \boldsymbol{\beta}_j$. The conditional posterior for $\boldsymbol{g}_j$ is:

$$p(\boldsymbol{g}_j = 1|\cdot) = \frac{1}{1 + \exp(-2\boldsymbol{\zeta}_j)}. \tag{23}$$

The conditional posteriors distributions for $\boldsymbol{\beta}_0$ and σ^2 are due to the conjugate nature of their priors a Gaussian and an inverse-Gamma distribution, respectively. They are given in their well-known form as:

$$p(\boldsymbol{\beta}_0|\cdot) \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\beta}_0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_0}\right)$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}_0} = \left(\frac{1}{\sigma^2}\mathbf{X}'_0\mathbf{X}_0 + \underline{\boldsymbol{\Sigma}}_{\beta_0}\right)^{-1}$$

$$\boldsymbol{\mu}_{\boldsymbol{\beta}_0} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}_0}\left(\frac{1}{\sigma^2}\mathbf{X}'_0\boldsymbol{y} + \underline{\boldsymbol{\Sigma}}_{\beta_0}\underline{\boldsymbol{\mu}}_{\beta_0}\right)$$

and

$$p(\sigma^2|\cdot) \sim \boldsymbol{\Gamma}^{-1}\left(N/2 + \underline{a}_{\sigma^2}, \ \underline{b}_{\sigma^2} + \frac{\sum_i^N (y_i - \eta_i)^2}{2}\right) \tag{24}$$

where η_i is the i-th element of the $N \times 1$ vector $\boldsymbol{\eta} = \boldsymbol{\eta}_0 + \sum_{j=1}^{Q} \mathbf{Z}_j \boldsymbol{\beta}_j$. Unfortunately the conditional distribution of ρ is of no well-known form and is given by:

$$p(\rho|\cdot) \propto |\mathbf{A}| \exp\left(-\frac{1}{2\sigma^2}(\mathbf{A}\boldsymbol{y} - \boldsymbol{\eta})'(\mathbf{A}\boldsymbol{y} - \boldsymbol{\eta})\right) p(\rho). \tag{25}$$

This implies that $p(\rho|\cdot)$ is not readily available, which prevents the usage of simple Gibbs steps for this parameter. ρ can, however, be sampled via a Metropolis-Hastings step.

Sampling procedure

With the conditional posteriors for $\boldsymbol{\alpha}$, γ_j, ν_j^2, ω, $\boldsymbol{\zeta}$, $\boldsymbol{g}$, $\boldsymbol{\beta}_0$, σ^2 and ρ given by Eqs. (19) - (25), respectively, MCMC estimation of the above outlined model involves sequential updating of the model parameters for T steps. We denote the value of the parameter at step t $(t = 0, \ldots, T)$ with the superscript (t). Thus, we can write the MCMC algorithm as:

(a) Initialise $\boldsymbol{\gamma}^{(0)}$, $\boldsymbol{\nu^2}^{(0)}$, ω, $\boldsymbol{g}^{(0)}$, $\boldsymbol{\beta}_0^{(0)}$, $\boldsymbol{\beta}^{(0)}$, $\sigma^{2(0)}$ and $\rho^{(0)}$ using maximum likelihood estimates or draws from the prior.

(b) Compute $\boldsymbol{\alpha}^{(0)}$, $\boldsymbol{\zeta}^{(0)}$ and $\mathbf{X}_\alpha^{(0)}$.

(c) Obtain a draw for $\boldsymbol{\alpha}^{(t)}$ via the conditional posterior $p(\boldsymbol{\alpha}^{(t)}|\cdot)$ in Eq. (19).

(d) Calculate $\mathbf{Z}_\zeta^{(t)} = \mathbf{Z} \operatorname{diag}\left(\operatorname{blockdiag}\left[\boldsymbol{\iota}_{l_1 1}, \ldots, \boldsymbol{\iota}_{l_Q Q}\right] \boldsymbol{\alpha}^{(t)}\right)$.

(e) Sample $\boldsymbol{g}_j^{(t)}$ from the posterior $p(\boldsymbol{g}_j^{(t)}|\cdot)$, see Eq. (23), for $j = 1, \ldots, Q$.

(f) Sample $\boldsymbol{\zeta}^{(t)}$ from the conditional posterior $p(\boldsymbol{\zeta}^{(t)}|\cdot)$, see Eq. (23).

(g) Rescale $\boldsymbol{\zeta}_j^{(t)}$ and $\alpha_j^{(t)}$ as in Eqs. (26) and (27), for $j = 1, \ldots, Q$.

(h) Calculate $\mathbf{Z}_\alpha^{(t)} = \mathbf{Z} \operatorname{blockdiag}\left(\boldsymbol{\zeta}_1^{(t)}, \ldots, \boldsymbol{\zeta}_Q^{(t)}\right)$.

(i) Update $\nu_1^{2(t)}, \ldots, \nu_Q^{2(t)}$ from their conditional posterior in Eq. (21).

(j) Update $\gamma_1^{2(t)}, \ldots, \gamma_Q^{2(t)}$ from their conditional posterior in Eq. (20).

(k) Draw $\omega^{(t)}$ from its consitional posterior $p(\omega^{(t)}|\cdot)$ as per Eq. (22).

(l) Update $\boldsymbol{\beta}_0^{(t)}$ from their conditional posterior, see Eq. (24).

(m) Update $\sigma^{2(t)}$ from its posterior in Eq. (24).

(n) Finally, use a random walk Metropolis step to sample $\rho^{(t)}$ with the following proposal density: $\mathcal{N}(\rho^{(t-1)}, \nu_\rho^2)$, where $\rho^{(t-1)}$ denotes the previous value and ν_ρ^2 governs the step size.

A fixed number of draws are stored after having discarded the first set of draws as burn-in (see, for example in Koop (2003)).

The rescaling in step (g) is suggested by Scheipl et al. (2012) as an important method to ensure identifiability for α_j and $\boldsymbol{\zeta}_j$. Without the rescaling, α_j and $\boldsymbol{\zeta}_j$ are not identifiable and their sampled values can wander to extreme regions of the posterior distribution, without a change in fit. This could for example happen if value of α_j are extremely large and sampled values of $\boldsymbol{\zeta}_j$ would be extremely slow. To avoid this pitfall, the following rescaling is applied:

$$\boldsymbol{\zeta}_j \rightarrow \frac{L_j}{\sum_{l_j}^{L_j} |\zeta_{jl_j}|} \boldsymbol{\zeta}_j \qquad (26)$$

$$\alpha_j \rightarrow \frac{\sum_{l_j}^{L_j} |\zeta_{jl_j}|}{L_j} \alpha_j \qquad (27)$$

The rescaling given by Eq. (26) and (27) ensures that α_j is scaled in a way that leaves $\boldsymbol{\beta}_j = \alpha_j \boldsymbol{\zeta}_j$ unchanged, but ensures that $|\boldsymbol{\zeta}_j|$ has mean of one. This completes the description of the Markov Chain Monte Carlo algorithm employed.

5   European freight generation and attraction

In this section we apply the spatial semi-parametric freight generation and attraction model to a cross-section of 258 regions. Thereby we illustrate the applicability of our approach for unveiling spatial structures, identifying model covariates and nonlinear influences of explanatory variables.

Data and space

Our data sample consists of cross-sectional observations on 258 European regions comprising 25 countries over the period of 2010-2014. Our unit of observation is the NUTS-2 level region, as defined by the 2010 revision of the European Commission. Although these regions correspond to administrative boundaries, they are frequently used in regional economics to model the sub-national variations prevalent in countries. Our sample consists of regions located in Austria (nine regions), Belgium (11 regions), Bulgaria (six regions), Czech Republic (eight regions), Germany (38 regions), Denmark (five regions), Estonia (one region), Greece (13 regions), Spain (15 regions), Finland (four regions), France (22 regions), Hungary (seven regions), Ireland (two regions), Italy (21 regions), Latvia (one region), Luxembourg (one region), Lithuania (one region), the Netherlands (12 regions), Poland (16 regions), Portugal (five regions), Romania (eight regions), Sweden (eight regions), Slovenia (two regions), Slovakia (four regions) and the United Kingdom (37 regions). For a complete list of the NUTS-2 regions included in this study, refer to Table B1 in Appendix B. This contains all EU-27 countries except Cyprus and Malta. The latter two have been excluded from the analysis due to the lack of available freight data from *Eurostat*.

Our dependent variables are the average total freight – measured in ten million tons – generated and attracted by NUTS-2 regions in the period of 2010 to 2014. The total generated and attracted freight was calculated as the sum of road, rail, inland waterways, maritime and air freight goods being loaded and unloaded in the study regions. All data tables on the amount of freight loaded and unloaded stem from the European Comission's (2016) *Eurostat* database. Fig. 1 shows on a map the NUTS-2 regions included in the study and the average amount of yearly generated and attracted freight.

[Fig. 1 about here.]

We consider a set of $K = 20$ explanatory variables, as well as their spatially lagged variants. In order to avoid potential problems with endogeneity, all explanatory variables are observed in 2006. Table 1 contains a list of the explanatory variables in the study, a detailed description of each variable, as well as an overview of their respective data sources.

Almost all freight generation and attraction studies include a variable measuring per capita income. We include regional domestic product per capita measured in millions of Euro in both of our models. In the freight generation model it serves as a measure for regional productivity, while in the freight attraction model it proxies regional per capita demand. Following Celik and Guldmann (2007), we also include the gross value added contribution of the manufacturing sector. They argue that this is a significant measure for the relative sectoral competitiveness of regions.

Furthermore, employment either in total, or differentiated by specific sectors, plays a significant role, both in explaining freight generation and freight attraction. In the case of freight generation, employment is an indication of the relative strength of a particular industry, and especially the manufacturing sector should emerge as a main driver of freight generation. In the case of freight attraction, sectoral employment in manufacturing and construction can be seen as a proxy for the demand for intermediate goods. Motivated by this, we include the share of employment in the manufacturing, construction and market services sectors in our analysis. Based on Novak et al. (2008), Chow et al. (2010) and Sánchez-Díaz et al. (2016) we expect the share of employment in manufacturing to be strongly significant. Novak et al. (2008) test different functional forms of sectoral employment and find strong support for quadratic influences.

We include further sectoral information in the form of the gross value contribution of regions' manufacturing. This sectoral indicator is expected to be strongly significant, as evidence presented by Chow et al. (2010) indicates that in both freight generation and attraction the manufacturing sector plays a central role. Furthermore, in the spirit of Chun et al. (2012) we include the average productivity of manufacturing companies, measured by the ratio of sectoral gross value added and the number of local units in the region. This aggregate explanatory variable serves as a measure for manufacturing competitiveness. Theory indicates that more competitive regions would generate and attract a greater volume of freight.

The characteristics of regional population play an important role in both freight attraction (Chun et al., 2012) and freight generation (Celik and Guldmann, 2007). In the case of freight attraction population density can be seen as a measure of consumer demand, in the case of freight generation, it is interpreted as a measure for urban and industrial concentration. Ranaiefar et al. (2013) find considerable evidence for the nonlinear influence of population density on freight attraction as well. Therefore, we follow Chow et al. (2010) and include population density as an explanatory variable in both models. As additional indication of a regions' population – especially to indicate the structure of urban centres – we add a set of dummy variables: namely, variables indicating whether a region contains a national capital, and major cities, and whether a region is predominantly rural. The findings on the influence of capital city regions are varied in literature. While a high population density seems to attract and generate a greater volume of freight, this does not seem to be true for major cities. Usually the largest factories and manufacturing centres are located at the outskirts of cities, therefore possible spatial spillovers might occur.

Multiple recent publications (Celik and Guldmann, 2007; Novak et al., 2008; Tavasszy et al., 2012; Wagner, 2010) emphasize the importance of logistics related factors for freight attraction and generation. In order to proxy these potential effects of logistics, we include the share of companies and the share of employment in land transport and pipelines, together with the warehousing activities per region. Moreover, we measure the number of commercial maritime ports and the number of commercial airports per region, as a proxy for transportation infrastructure, in combination with the length of the road network and the closest travel time to maritime ports. Almost all studies show the importance of transport infrastructure for freight generation and attraction.

[Table 1 about here.]

Moreover, we follow Lawson et al. (2012), Sánchez-Díaz et al. (2016) and Chow et al. (2010), and include urbanization – the share of urban areas per the population density of the region – as a measure of land-use activities. Higher urbanization typically indicates a higher rate of manufacturing and production activities and larger demand markets.

Finally, we have included some indicators of our own; by measuring the effects of national borders, whether a region has received Objective 1 funding in the period 2000 to 2006, and whether a region has a sea coast (see Table 1). Table B2 in Appendix B contains summary statistics of the dependent and independent variables used in the study.

For constructing the spatial neighborhood matrix we use the geodesic distance between the centroids of NUTS-2 regions included in our analysis. We tested model specification containing a spatial neighborhood matrix with five to twelve nearest neighbors. The results did not change significantly. The current results all stem from a model with a spatial weight matrix with a seven nearest neighbor specification. We used Moran's I as a measure of spatial correlation prevalent in the data. The test yielded a Moran's I value of $0,1596$ ($p < 0.001$) for freight generation and $0,1417$ ($p < 0.001$) in the case of freight attraction.

Prior elicitation and computational notes

For both the freight generation and attraction models we use the same prior set-up. For the priors of the Normal inverse-Gamma hyperprior structure we set the overall prior inclusion probability of variables as $\underline{a}_\omega = 1$ and $\underline{b}_\omega = 1$. This is the most agnostic set-up, since a priori all semiparametric model components have an inclusion probability of 0.5, and it corresponds to a uniform prior for ω. For the prior variance of α_j, we set $\underline{a}_\nu = 1$ and $\underline{b}_\nu = 25$, which is a rather non-informative setting for the prior variances. The shrinkage parameter is set to $\kappa_0 = 10^{-6}$.

For the parametric part of our model, we again try to be as noninformative as possible in our selection of priors. Therefore, we set the prior for the overall model variance as $\underline{a}_{\sigma^2} = 0.001$ and $\underline{b}_{\sigma^2} = 0.001$. Only the intercept is included in $\mathbf{X}_0$, thus $K_0 = 1$. For the prior mean and prior variance of the intercept we use diffuse distributions, setting the prior mean to zero and $\underline{\Sigma}_{\mu_0} = 10^4 \mathbf{I}_{K_0}$.

We set the number of spline knots as $L_m = 20$. The position of spline knots is equally spaced along along the range of each covariate. Over these spline knots we construct cubic splines, with a first order penalization matrix. Cubic splines should adequately approximate potential nonlinearities and setting a penalization matrix of the first degree mean that the unpenalized part directly corresponds to the linear coefficient vector, which lies in the nullspace of the penalization matrix.

Estimation results

Posterior inference is based on 20,000 draws with the first 10,000 discarded as burn-ins.[4] Table 2 and Table 3 report the results of the analysis for the freight generation and attraction models, respectively. The rows in both tables correspond to all $K = 24$ explanatory variables and their spatially lagged variants. Column (a) contains the posterior inclusion probability of the unpenalized variable components. An inclusion probability of one corresponds to the variable being included in all the sampled models and an inclusion probability of zero indicates that the variable was excluded from all sampled models. Columns (b) to (e) contain the posterior mean, posterior standard deviation and posterior sign certainty of the unpenalized regression coefficients β_j. Column (e) contains the posterior inclusion probability of the penalized model components, that is the posterior probability $\gamma_j = 1$ for each of the covariates modeled in a semi-parametric fashion. If the unpenalized posterior inclusion probability of a variable [in column (a)] is close to one and the penalized posterior inclusion probability [column (e)] is close to zero, this can be interpreted as the variable having a predominantly linear impact. If the penalized inclusion probability is close to one, but the penalized part is close to zero, this means that the mean impact of the covariate on the dependent variable is zero, but some nonlinear influences are prevalent. If both the inclusion probability in column (a) and column (e) display an inclusion probability close to one, this can be interpreted as the variable having a strongly nonlinear influence on freight generation or attraction.

Finally, column (f) in Table 2 and Table 3 contains a graphical representation the posterior functional forms of the penalized regression coefficients. This column can be interpreted similarly to its mean, standard deviation and sign probability counterparts in columns (b) - (e) and is

---

[4]We assessed the convergence of our sampler using the diagnostic suggested by Geweke (1992) and implemented in the R-package `coda`. The obtained z-scores of 0.84 and 0.88 for freight generation and attraction, respectively, suggest a successful convergence of the sampler.

used as a practical method of summarizing 20 regression coefficients corresponding to $\boldsymbol{\beta}_j$. Each graphic shows the posterior impact of $f_j(\cdot)$ on $\boldsymbol{y}$ in the interval ± 1. The shaded areas confer to 80% confidence intervals and the continuous black line corresponds to the zero-line. It is readily observable that all functions with a penalized posterior inclusion probability [column (e)] close to zero are strongly shrunk to a straight line at $\boldsymbol{y} = 0$, while function with inclusion probability close to one display significant nonlinear influence of the j-th co-variate.

[Table 2 about here.]

The posterior inclusion probabilities of the co-variates in the freight generation model are presented in Table 2. The posterior inclusion probabilities of the explanatory variables display a similar pattern to the freight attraction model in Table 3, with small differences in the specific coefficient estimates. Both *Manufacturing gross value added* and its spatially lagged counterpart have posterior inclusion probability of unity, that is they have been included in all posterior models. Only the non-spatially lagged semi-parametric model term of *Manufacturing gross value added* exhibits a posterior inclusion probability of close to unity. Moreover, the penalized model term of **W** *Population density* exhibits a high inclusion probability, close to one, while its linear version is excluded from the majority of the posterior models ($\gamma_j = 0.0537$). A region being a capital city seems to play a central role in the majority of the models ($\gamma_j = 0.9825$) with a sign probability of unity towards a negative impact. The spatial coefficient ρ is highly significant with $\rho = 0.4274$ (s.d. 0.0670).

[Table 3 about here.]

Turning our attention to the posterior inclusion probabilities in the freight attraction model (Table 3), we can see that the spatial parameter ρ is significant at 0.2485 (s.d. 0.0996) and positive, which indicates the presence of positive spatial spillovers. This is inline with earlier findings by Novak et al. (2008), who also found strongly significant and positive spatial autocorrelation. Furthermore, we observe that the *Manufacturing gross value added* and its spatially lagged counterpart exhibit an inclusion probability close to one. This is in-line with similar findings from Chow et al. (2010), where the intermediate demand markets (proxied by the gross value added in manufacturing activities) play the central role in attracting the majority freight, as opposed to consumer demand markets. The differing signs for the coefficient *Manufacturing gross value added* and its spatially lagged counterpart indicate some support for Ranaiefar et al. (2013), where logistical activity is strong on regions neighboring manufacturing sites. Moreover, the penalized component of *Manufacturing gross value added* also has a high posterior inclusion probability. The functional form in column (f) indicates that higher values of *Manufacturing gross value added* have a proportionally stronger impact on attracting freight than lower values.

While in their unpenalized form the coefficients corresponding to *Population density* and its spatially lagged counterpart do not exhibit posterior inclusion probabilities above 0.9, the penalized model component corresponding to **W** *Population density* is included in all posterior models. The functional form depicted in column (f) indicate that population density up to 1,000 inhabitants per square km has a comparatively negative effect, while population density in the ranges from 1,000 inhabitants to 1,600 inhabitants to square km seem to have a comparatively positive effects on freight attraction. The covariate *Capital city region* also exhibits a posterior inclusion probability close to one.

[Table 4 about here.]

For the purposes of comparing the impact of spatial dependence on the model, we ran corresponding freight generation and attraction models without spatial dependence, that is $\rho = 0$. The results are displayed in Table 4.[5] The residuals of the freight generation and attraction models both display significant spatial dependence, with a Moran's I of 0.142 and 0.141 (and both with $p_¡0.001$),

---

[5]Our posterior inference is based again on 20,000 draws, with the first 10,000 discarded as burn-ins. Results obrained from the convergence diagnostics of Geweke (1992), as implemented in the R-package `coda`, suggest a successful convergence of the sampler, with test values of 0.86 and 0.91 for freight generation and attraction, respectively.

respectively. We compare both models to their spatially lagged counterparts using Bayes factor and a non-informative prior (see Koop 2003). In both cases the test favours the model containing the spatial lags. The resulting parameter estimates also show considerable bias, both in the freight generation and attraction models. As a further point of comparison, Table B3 in Appendix B presents the estimated results from a non-semi-parametric spatial Durbin model. The model was run using the MCMC sampler presented in LeSage and Pace (2009), Chapter 5, Section 3.

Figure 2 displays the functional fit for the two penalized model terms with the highest posterior inclusion probability, for the freight generation and freight attraction model, respectively. The posterior mean functional form of *Manufacturing gross value added* exhibits initially low impacts in the range of 0.01 - 0.068, both in the freight attraction and generation models – see panels (i) and (iv). In the range of 1.11 - 1.32 the posterior mean of *Manufacturing gross value added* shows comparatively higher positive impacts on freight attraction and generation, respectively. This indicate that the impact of *Manufacturing gross value added* does not continuously remain the same over its range of values, but increases sharply.

The functional form of the penalized counterpart of the model term **W** *Population density* in panels (ii) and (iii) exhibits a positive spike at 1.34 and 1.37, respectively. Otherwise, it is negative or not significantly different from zero. This indicates that a region with a population density of 1.340 – 1.370 inhabitants per square km is much more likely to generate or attract freight. This is in all likelihood a proxy effect for industrial and/or warehousing regions, which share a specific pattern of population density.

[Fig. 2 about here.]

Table 5 shows the posterior spatial impact estimates of both the freight generation – displayed in panel (a) – and freight attraction model – shown in panel (b). The first three columns correspond to the posterior mean, standard deviation and sign probability of the average direct impacts, respectively. Columns 4-6 display the corresponding values for the indirect impacts and columns 7-9 show the posterior total impact coefficients.

In both models *Manufacturing gross value added* has a sign probability of close to unity and is positive. According to these results, if in a region the gross value added contribution of the manufacturing sector increases by one million Euro then, freight generation in that region will increase by 2.2 million tons (std. dev. 0.519) and freight attraction by 1.9 million (std. dev. 0.485). Conversely, an average increase in the gross value added contribution of the manufacturing sector by one million euro in a regions' neighbors, should increase freight attraction by 1.7 million tons (std. dev. 0.465) and freight generation by 1.4 million (std. dev. 0.405). If a region is a capital city, this decreases its freight generation in comparison to non-capital city regions by -0.49 million tons and freight attraction by -0.48 million tons. In regions neighboring capital cities, freight attraction is decreased by -0.36 million tons and freight generation by -0.35 million tons, on average.

[Table 5 about here.]

In conclusion we note that our approach allowed us to provide estimates and inference, regarding which variables out of a set of 20 contribute significantly to freight generation and attraction. Moreover, we could infer that the manufacturing sector seems to have a nonlinear in the parameters impact on both freight generation and attraction and we could provide an estimate of the functional form. Furthermore, spatial spillovers are significant in both models at the values 0.427 and 0.248, respectively.

6   Closing remarks

The classic regional freight generation and attraction models suffer from a series of weaknesses. First, they do not take spatial lags of the explanatory or dependent variables into account, thus leading to biased and inconsistent estimates (Novak et al., 2008). Second, there is no clear consensus as to the choice of explanatory variables in the models (Chow et al., 2010). The classical model emphasizes gross regional product, population density and indicators of sectoral performance,

several authors argue for land-use indicators (Boerkamps et al. 2000,Chow et al. 2010, Wagner 2010 and Wisetjindawat et al. 2006) or indicators of regional logistic and warehousing activities (Hesse and Rodrigue 2004, Tavasszy et al. 2012). Third, there is considerable evidence for nonlinearity in the models, however, there is no clear consensus on the functional form of this nonlinearity (Rodrigue 2006, Hesse and Rodrigue 2004 and Ranaiefar et al. 2013).

Novak et al. (2008) provide evidence for siginficant spatial spillovers in both freight attraction and freight generation. Based on semi-parametric methods and in the spirit of Scheipl et al. (2012), we use an approach based on penalized basic splines to approximate possible nonlinearities. Moreover, we implement Bayesian variable selection through a penalized Normal inverse-Gamma prior structure, which enables us to make inference over the coefficients with the highest posterior inclusion probability. Furthermore, this prior structure, coupled with the penalized spline representation, enables inference over the nonlinear influence of variables.

In the context of European freight generation, we find considerable influence for spatial dependencies among both the freight generation and attraction of NUTS-2 level regions. Furthermore, we provide evidence that the manufacturing sector seems to play a key role in European freight generation, with an increase in the gross value added resulting in a significant increase in freight generation. Finally, we show that both the gross value added of the manufacturing sector, as well as the population of neighbouring regions exhibit a nonlinear influence on freight generation and attraction. This provides a clear evidence for the nonlinearities stemming from logistical and transportation related factors, as advocated by Rodrigue (2006) and Ranaiefar et al. (2013).

References

Anselin L (1988) *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht

Anselin L and Bera A (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah A and Giles DEA, eds., *Handbook of Applied Economic Statistics*. Marcel Dekker, New York, 237–289

Arbia G (2006) *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*. Springer Science & Business, Berlin Heidelberg New York

Basile R (2008) Regional economic growth in Europe: A semiparametric spatial dependence approach. *Papers in Regional Science* 87(4), 527–544

Basile R, Mínguez R, Montero JM and Mur J (2014) Modelling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities. *Journal of Economic Dynamics and Control* 48(1), 229–245

Boerkamps J, Binsbergen AV and Bovy P (2000) Modeling behavioral aspects of urban freight movement in supply chains. *Transportation Research Record: Journal of the Transportation Research Board* 1725(1), 1–11

Brezger A and Lang S (2006) Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics & Data Analysis* 50(4), 967–991

Celik MH and Guldmann JM (2007) Spatial interaction modeling of interregional commodity flows. *Socio-Economic Planning Sciences* 41(2), 147–162

Chow JYJ, Yang CH and Regan AC (2010) State-of-the art of freight forecast modeling: Lessons learned and the road ahead. *Transportation* 37(6), 1011–1030

Chun Y, Kim H and Kim C (2012) Modeling interregional commodity flows with incorporating network autocorrelation in spatial interaction models: An application of the US interstate commodity flows. *Computers, Environment and Urban Systems* 36(6), 583–591

De Grange L, Boyce D, González F and Ortúzar JDD (2013) Integration of spatial correlation into a combined travel model with hierarchical levels. *Spatial Economic Analysis* 8(1), 71–91

De Grange L, Fernández E and De Cea J (2010) A consolidated model of trip distribution. *Transportation Research Part E: Logistics and Transportation Review* 46(1), 61–75

De Jong G, Vierth I, Tavasszy L and Ben-Akiva M (2013) Recent developments in national and international freight transport models within Europe. *Transportation* 40(2), 347–371

DeBoor C (1978) *A Practical Guide to Splines*. Springer, Berlin Heidelberg New York

Del Bo CF and Florio M (2012) Infrastructure and growth in a spatial framework: Evidence from the EU regions. *European Planning Studies* 20(8), 1393–1414

Eilers PHC and Marx BD (1996) Flexible smoothing with B-splines and penalties. *Statistical Science* 11(2), 89–121

Fahrmeir L, Kneib T and Lang S (2004) Penalized structured additive regression for space-time data: A Bayesian perspective. *Statistica Sinica* 14(3), 715–745

Fischer MM and Wang J (2011) *Spatial Data Analysis: Models, Methods and Techniques*. Springer, Heidelberg Dordrecht London New York

Fotopoulos G (2012) Nonlinearities in regional economic growth and convergence: The role of entrepreneurship in the European union regions. *Annals of Regional Science* 48(3), 719–741

George EI and McCulloch RE (1993) Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 88(423), 881–889

George EI and McCulloch RE (1997) Approaches for Bayesian variable selection. *Statistica Sinica* 7(2), 339–373

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo JM, Berger JO, Dawid AP and Smith AFM, eds., *Bayesian Statistics*. Oxford University Press, Oxford, 4. edition, 167–193

Hesse M and Rodrigue JP (2004) The transport geography of logistics and freight distribution. *Journal of Transport Geography* 12(3), 171–184

Ishwaran H and Rao JS (2005) Spike and slab variable selection: Frequentist and Bayesian strategies. *Annals of Statistics* 33(2), 730–773

Koop G (2003) *Bayesian Econometrics*. John Wiley & Sons Ltd., West Sussex

Koop G and Poirier D (2004) Bayesian variants of some classical semiparametric regression techniques. *Journal of Econometrics* 123(2), 259–282

Kuo L and Mallick B (1998) Variable selection for regression models. *Sankhya: The Indian Journal of Statistics* 60(1), 65–81

Lang S and Brezger A (2004) Bayesian P-Splines. *Journal of Computational and Graphical Statistics* 13(1), 183–212

Lawson C, Holguín-Veras J, Sánchez-Díaz I, Jaller M, Campbell S and Powers E (2012) Estimated generation of freight trips based on land use. *Transportation Research Record: Journal of the Transportation Research Board* 2269(1), 65–72

Lesage JP and Fischer MM (2009) Spatial growth regressions: Model specification, estimation and interpretation. *Spatial Economic Analysis* 3(3), 275–304

LeSage JP and Parent O (2007) Bayesian model averaging for spatial econometric models. *Geographic Analysis* 39(3), 241–267

LeSage PJ and Pace RK (2009) *Introduction to Spatial Econometrics.* CRC Press, Boca Raton London New York

Novak DC, Hodgdon C, Guo F and Aultman-Hall L (2008) Nationwide freight generation models: A spatial regression approach. *Networks and Spatial Economics* 11(1), 23–41

Ortúzar JdD and Willumsen LG (2011) *Modelling Transport.* John Wiley & Sons, Chichester, 4th edition

Piribauer P and Cuaresma JC (2016) Bayesian variable selection in spatial autoregressive models. *Spatial Economic Analysis*

Piribauer P and Fischer MM (2015) Model uncertainty in matrix exponential spatial growth regression models. *Geographical Analysis* 47(3), 240–261

Ranaiefar F, Chow JYJ, Rodriguez-Roman D, Veiga de Camargo P and Ritchie SG (2013) Geographic scalability and supply chain elasticity of a structural commodity generation model using public data. Technical report, Institute of Transportation Studies, University of California

Rodrigue JP (2006) Challenging the derived transport-demand thesis: Geographical issues in freight distribution. *Environment and Planning A* 38(8), 1449–1462

Ruppert D, Wand MP and Carroll RJ (2003) *Semiparametric Regression.* Cambridge University Press, Cambridge

Sánchez-Díaz I, Holguín-Veras J and Wang X (2016) An exploratory analysis of spatial effects on freight trip attraction. *Transportation* 43(1), 177–196

Scheipl F, Fahrmeir L and Kneib T (2012) Spike-and-slab priors for function selection in structured additive regression models. *Journal of the American Statistical Association* 107(500), 1518–1532

Tavasszy LA, Ruijgrok K and Davydenko I (2012) Incorporating logistics in freight transport demand models: State-of-the-art and research opportunities. *Transport Reviews* 32(2), 203–219

Wagner T (2010) Regional traffic impacts of logistics-related land use. *Transport Policy* 17(4), 224–229

Wisetjindawat W and Sano K (2003) A behavioral modeling in micro-simulation for urban freight transportation. *Journal of the Eastern Asia Society for Transportation Studies* 5(3), 2193–2208

Wisetjindawat W, Sano K and Matsumoto S (2006) Commodity distribution model incorporating spatial interactions for urban freight movement. *Transportation Research Record: Journal of the Transportation Research Board* 1966(1), 41–50

Table 1: Variables used in the analysis

Variable	Description
Freight generated	Annual road, marine, inland waterway, rail and air freight transport by region of loading (average 2011-2014); measured in ten million tons. (Source: *Eurostat*)
Freight attracted	Annual road, marine, inland waterway, rail and air freight transport by region of unloading (average 2011-2014); measured in ten million tons. (Source: *Eurostat*)
Regional domestic product per capita	Regional domestic product per inhabitant, in 100,000 Euro, 2006. (Source: *Eurostat*)
Manufacturing gross value added	Gross value added of NACE rev 2 C to I (manufacturing) in billions of Euro, 2006. (Source: *Eurostat*)
Population density	Share of 1.000 inhabitants per square km, 2006. (Source: *Eurostat*)
Manufacturing employment	Share of NACE rev 2 C to I (manufacturing) in total employment, 2006. (Source: *Eurostat*)
Construction employment	Share of NACE rev 2 F (construction) in total employment, 2006.
Market services employment	Share of NACE rev 2 G to K (market services) in total employment, 2006. (Source: *Eurostat*)
Avg. production of manufacturing units	Gross value added per average number of employees of NACE rev 2 B to E (manufacturing) companies, in EUR, 2006. (Source: *Eurostat*)
Degree of urbanization	Share of urban surfaces (CLC 11) per population density, 2006. (Source: *Eurostat*)
Share of logistic companies	Share of local units of NACE rev 2 sectors H49 (Land transport and transport via pipelines) and H52 (Warehousing and support activities for transportation) from total local companies, 2006. (Source: *Eurostat*)
Logistic employment (share)	Share of employees in NACE rev 2 sectors H49 (Land transport and transport via pipelines) and H52 (Warehousing and support activities for transportation), 2006. (Source: *Eurostat*)
Number of airports	Number of commercial airports open to all flights in the region, 2006. (Source: *Eurostat*)
Length of road network (km)	Length of road network in 10,000 km, 2001 (Source: *ESPON*)
Number of seaports	Number of major commercial maritime ports in the region, 2001. (Source: *Eurostat*)
Travel time to seaport (min)	Car travel time in minutes from region's centroid to the closest major maritime port, 2001. (Source: *Eurostat*)
Capital city region	Dummy variable; 1 denotes the presence of a national capital in the region, 0 otherwise (Source: *ESPON*)
Region with large city	Dummy variable; 1 denotes the presence of a city (inhabitants > 100,000) in the region, 0 otherwise (Source: *Eurostat*)
Rural region	Dummy variable; 1 denotes a region with population density below 100 and containing no city with more than 125,000 inhabitants (Source: *ESPON*)
Border region	Dummy variable; 1 denotes a region with a national border, 0 otherwise (Source: *Eurostat*)
Objective 1 region	Dummy variable; 1 indicates that the region was eligible to apply for Objective 1 funding in 2000-2006, 0 otherwise. (Source: *ESPON*)
Coastal region	Dummy variable; 1 denotes that the region has a marine coast, 0 otherwise (Source: *ESPON*)

Table 2: Freight generation model posterior estimates

Variable	Unpenalized model part				Penalized model part	
	Inclusion prob. (a)	Mean (b)	Std. dev. (c)	Sign prob. (d)	Inclusion prob. (e)	functional form (f)
ρ		0.4274	0.0670	1.0000		
σ^2		0.1455	0.0311			
Regional domestic product per capita	0.0342	0.0070	0.0612	0.5217	0.0812	
W Regional domestic product per capita	0.0223	0.0043	0.0412	0.5221	0.0384	
Manufacturing gross value added	1.0000	-2.6829	0.3251	1.0000	0.9284	
W Manufacturing gross value added	1.0000	1.0762	0.2716	0.9998	0.0104	
Population density	0.0836	0.0272	0.1023	0.5500	0.0206	
W Population density	0.0537	0.0379	0.2366	0.5158	0.9946	
Manufacturing employment	0.0147	0.0003	0.0246	0.5013	0.0146	
W Manufacturing employment	0.0225	0.0015	0.0316	0.5102	0.0087	
Construction employment	0.0171	0.0004	0.0309	0.5074	0.0159	
W Construction employment	0.0329	-0.0021	0.0465	0.5040	0.0103	
Market services employment	0.0619	-0.0150	0.0777	0.5267	0.0050	
W Market services employment	0.0118	-0.0004	0.0219	0.5107	0.0108	
Avg. production of manufacturing units	0.0254	-0.0036	0.0433	0.5107	0.5883	
W Avg. production of manufacturing units	0.0196	-0.0014	0.0289	0.5065	0.0972	
Degree of urbanization	0.0194	-0.0008	0.0173	0.5072	0.0165	
W Degree of urbanization	0.0136	-0.0027	0.0268	0.5201	0.0101	
Share of logistic companies	0.0238	-0.0034	0.0528	0.5084	0.1485	
W Share of logistic companies	0.0424	-0.0135	0.0797	0.5372	0.0225	
Logistic employment (share)	0.0412	-0.0129	0.0770	0.5189	0.0393	
W Logistic employment (share)	0.0193	-0.0022	0.0382	0.5144	0.0280	
Number of airports	0.1370	-0.0687	0.1981	0.5756	0.0175	
W Number of airports	0.0247	-0.0029	0.0358	0.5179	0.0112	
Length of road network (km)	0.0381	-0.0101	0.0629	0.5232	0.0627	
W Length of road network (km)	0.0192	0.0004	0.0266	0.5067	0.0595	
Number of seaports	0.0124	-0.0005	0.0210	0.5049	0.0133	
W Number of seaports	0.0278	0.0052	0.0473	0.5136	0.0119	
Travel time to seaport (min)	0.0122	-0.0004	0.0172	0.5025	0.0040	
W Travel time to seaport (min)	0.0563	-0.0267	0.1199	0.5474	0.0536	
Capital city region	0.9825	-0.3447	0.0755	1.0000		
W Capital city region	0.0159	-0.0023	0.0253	0.5175		
Region with large city	0.0334	0.0073	0.0299	0.5748		
W Region with large city	0.0533	0.0161	0.0688	0.5498		
Rural region	0.0136	-0.0020	0.0129	0.5358		
W Rural region	0.0206	0.0019	0.0214	0.5218		
Border region	0.0284	0.0059	0.0240	0.5670		
W Border region	0.1965	0.0568	0.1115	0.6547		
Objective 1 region	0.0091	0.0025	0.0145	0.5369		
W Objective 1 region	0.1251	0.0352	0.0844	0.6209		
Coastal region	0.0066	0.0001	0.0095	0.5008		
W Coastal region	0.0300	-0.0078	0.0392	0.5512		

Notes: Posterior results based on **W** matrix with seven nearest neighbors. The plots in column (f) are all bounded on the y-axis at ± 1. The shaded areas are 80% confidence interval, the dotted line represents the posterior mean of the spline function and the continuous line marks $y = 0$.

Table 3: Freight attraction model posterior estimates

Variable	Unpenalized model part				Penalized model part	
	Inclusion prob. (a)	Mean (b)	Std. dev. c	Sign prob. (d)	Inclusion prob. (e)	functional form (f)
ρ		0.2485	0.0996	1.0000		
σ^2		0.1338	0.0191			
Regional domestic product per capita	0.0504	0.0206	0.1121	0.5262	0.0358	
W Regional domestic product per capita	0.0306	0.0063	0.0508	0.5306	0.0463	
Manufacturing gross value added	1.0000	-2.9145	0.3344	1.0000	1.0000	
W Manufacturing gross value added	0.9555	1.0559	0.3658	0.9775	0.0279	
Population density	0.0279	0.0043	0.0386	0.5138	0.0283	
W Population density	0.1095	0.0688	0.3133	0.5342	1.0000	
Manufacturing employment	0.0122	0.0016	0.0307	0.5022	0.0136	
W Manufacturing employment	0.0648	0.0324	0.1498	0.5334	0.0099	
Construction employment	0.0154	-0.0001	0.0299	0.5048	0.0175	
W Construction employment	0.0134	-0.0013	0.0365	0.5049	0.0162	
Market services employment	0.0251	-0.0082	0.0599	0.5317	0.0097	
W Market services employment	0.0229	0.0011	0.0365	0.5032	0.0105	
Avg. production of manufacturing units	0.0379	-0.0089	0.0667	0.5171	0.5775	
W Avg. production of manufacturing units	0.0250	0.0003	0.0367	0.5045	0.1162	
Degree of urbanization	0.0164	-0.0013	0.0172	0.5090	0.0091	
W Degree of urbanization	0.0215	-0.0035	0.0310	0.5236	0.0090	
Share of logistic companies	0.0200	-0.0007	0.0616	0.5111	0.1157	
W Share of logistic companies	0.0465	-0.0158	0.0860	0.5287	0.0103	
Logistic employment (share)	0.0312	-0.0068	0.0591	0.5282	0.0275	
W Logistic employment (share)	0.0388	-0.0081	0.0639	0.5111	0.0261	
Number of airports	0.0817	-0.0336	0.1334	0.5539	0.0187	
W Number of airports	0.0290	-0.0034	0.0384	0.5131	0.0079	
Length of road network (km)	0.0405	-0.008	0.0546	0.5348	0.0307	
W Length of road network (km)	0.0125	-0.0003	0.0268	0.5027	0.0346	
Number of seaports	0.0128	-0.0027	0.0322	0.5110	0.0113	
W Number of seaports	0.0392	0.0089	0.0644	0.5137	0.0119	
Travel time to seaport (min)	0.0102	-0.0007	0.0154	0.5090	0.0079	
W Travel time to seaport (min)	0.0771	-0.0317	0.1226	0.5533	0.0429	
Capital city region	1.0000	-0.3645	0.0764	1.0000		
W Capital city region	0.0179	-0.0028	0.0282	0.5141		
Region with large city	0.0265	0.0061	0.0276	0.5631		
W Region with large city	0.0330	0.0092	0.0489	0.5444		
Rural region	0.0056	-0.0016	0.0115	0.5360		
W Rural region	0.0132	0.0012	0.0184	0.5004		
Border region	0.0250	0.005	0.0235	0.5580		
W Border region	0.3560	0.1034	0.1427	0.7204		
Objective 1 region	0.0136	0.003	0.0158	0.5521		
W Objective 1 region	0.0835	0.0259	0.0751	0.6019		
Coastal region	0.0069	0.0018	0.0130	0.5313		
W Coastal region	0.0271	-0.0058	0.0324	0.5428		

Notes: Posterior results based on **W** matrix with seven nearest neighbors. The plots in column (f) are all bounded on the y-axis at ± 1. The shaded areas are 80% confidence interval, the dotted line represents the posterior mean of the spline function and the continuous line marks $y = 0$.

Table 4: Non-spatial freight generation (a) and attraction (b) model posterior estimates

(a) - Freight Generation

Variable	Unpenalized model part				Penalized model part	
	Inclusion prob. (a)	Mean (b)	Std. dev. (c)	Sign prob. (d)	Inclusion prob. (e)	functional form (f)
Moran's I		0.1421	0.0301			
σ^2		0.1615	0.0178			
Regional domestic product per capita	0.1644	0.0262	0.1247	0.5451	0.1288	
Manufacturing gross value added	1.0000	-2.4002	0.3339	1.0000	0.9939	
Population density	0.4261	0.1868	0.2690	0.7096	0.1842	
Manufacturing employment	0.7092	0.3865	0.3187	0.8511	0.0999	
Construction employment	0.1127	0.0010	0.0832	0.5060	0.1126	
Market services employment	0.2258	-0.0794	0.1907	0.6058	0.0759	
Avg. production of manufacturing units	0.1637	-0.0287	0.1452	0.5355	0.9873	
Degree of urbanization	0.1016	-0.0061	0.0413	0.5289	0.0647	
Share of logistic companies	0.1054	-0.0042	0.0829	0.5147	0.1663	
Logistic employment (share)	0.1386	-0.0183	0.1093	0.5286	0.1224	
Number of airports	0.2341	-0.0746	0.1934	0.5928	0.0995	
Length of road network (km)	0.1671	-0.0225	0.1017	0.5516	0.2583	
Number of seaports	0.0988	0.0012	0.0545	0.5066	0.0664	
Travel time to seaport (min)	0.0766	-0.0011	0.0313	0.5009	0.0595	
Capital city region	0.6322	-0.1312	0.1138	0.8463		
Region with large city	0.1810	0.0245	0.0578	0.6171		
Rural region	0.0803	-0.0073	0.0296	0.5588		
Border region	0.5173	0.0764	0.0792	0.8017		
Objective 1 region	0.0521	0.0031	0.0180	0.5426		
Coastal region	0.0381	0.0010	0.0134	0.5107		

(b) - Freight Attraction

Variable	Unpenalized model part				Penalized model part	
	Inclusion prob. (a)	Mean (b)	Std. dev. (c)	Sign prob. (d)	Inclusion prob. (e)	functional form (f)
Moran's I		0.1415	0.0302			
σ^2		0.1728	0.0197			
Regional domestic product per capita	0.1256	0.0120	0.0940	0.5136	0.0944	
Manufacturing gross value added	1.0000	-2.6084	0.3482	1.0000	1.0000	
Population density	0.2522	0.0768	0.1800	0.5970	0.2372	
Manufacturing employment	0.6416	0.4051	0.3549	0.8335	0.1242	
Construction employment	0.1566	-0.0154	0.1172	0.5153	0.1077	
Market services employment	0.3242	-0.1280	0.2423	0.6459	0.0710	
Avg. production of manufacturing units	0.1385	-0.0228	0.1380	0.5387	0.9744	
Degree of urbanization	0.0929	-0.0065	0.0533	0.5303	0.0862	
Share of logistic companies	0.1344	-0.0116	0.1055	0.5224	0.1548	
Logistic employment (share)	0.1415	-0.0157	0.1078	0.5248	0.1538	
Number of airports	0.2341	-0.0522	0.1626	0.5720	0.0749	
Length of road network (km)	0.1255	-0.0142	0.0839	0.5318	0.2073	
Number of seaports	0.1153	-0.0075	0.0696	0.5193	0.0531	
Travel time to seaport (min)	0.0961	-0.0029	0.0358	0.5116	0.0572	
Capital city region	0.7972	-0.1889	0.1109	0.9257		
Region with large city	0.1413	0.0202	0.0539	0.6062		
Rural region	0.0629	-0.0045	0.0241	0.5474		
Border region	0.4177	0.0638	0.0798	0.7660		
Objective 1 region	0.0515	0.0023	0.0161	0.5273		
Coastal region	0.0652	0.0043	0.0219	0.5509		

Notes: The plots in column (f) are all bounded on the y-axis at ± 1. The shaded areas are 80% confidence interval, the dotted line represents the posterior mean of the spline function and the continuous line marks $y = 0$.

Table 5: Freight generation (a) and attraction (b) model spatial impact estimates

(a) - Freight generation

Variable	Average direct impacts			Average indirect impacts			Average total impacts		
	Mean	Std. dev.	Sign prob.	Mean	Std. dev.	Sign prob.	Mean	Std. dev.	Sign prob.
Regional domestic product per capita	-0.0852	0.1857	0.6798	-0.0630	0.1363	0.6796	-0.0222	0.0543	0.6794
Manufacturing gross value added	2.2269	0.5190	0.9998	1.6976	0.4646	0.9998	0.5293	0.2437	0.9998
Population density	0.2621	0.2854	0.8355	0.1923	0.2214	0.8208	0.0698	0.0789	0.8697
Manufacturing employment	-0.3558	1.4643	0.5639	-0.2944	1.2086	0.5636	-0.0614	0.3027	0.5634
Construction employment	0.0285	0.3648	0.5302	0.0214	0.2794	0.5301	0.0071	0.0960	0.5288
Market services employment	0.0314	0.2738	0.5335	0.0240	0.1998	0.5326	0.0074	0.0808	0.5354
Avg. production of manufacturing units	-0.3835	0.2692	0.9179	-0.2938	0.2056	0.9179	-0.0897	0.0798	0.9177
Degree of urbanization	-0.0040	0.0324	0.5524	-0.0031	0.0246	0.5521	-0.0009	0.0089	0.5527
Share of logistic companies	0.5666	1.3429	0.6407	0.4429	1.0391	0.6416	0.1237	0.3389	0.6381
Logistic employment (share)	0.6792	2.5502	0.5723	0.5095	1.9088	0.5740	0.1696	0.7102	0.5695
Number of airports	0.0033	0.0164	0.5704	0.0024	0.0121	0.5712	0.0009	0.0049	0.5698
Length of road network (km)	0.1433	0.3516	0.6682	0.1050	0.2539	0.6683	0.0383	0.1090	0.6686
Number of seaports	-0.0015	0.0084	0.5392	-0.0011	0.0060	0.5393	-0.0004	0.0027	0.5372
Travel time to seaport (min)	0.0318	0.0753	0.6457	0.0241	0.0564	0.6460	0.0077	0.0212	0.6457
Capital city region	-0.4979	0.1309	1.0000	-0.3682	0.0772	1.0000	-0.1296	0.0812	0.9965
Region with large city	0.0197	0.0710	0.5928	0.0064	0.0278	0.5815	0.0133	0.0614	0.5884
Rural region	-0.0007	0.0285	0.5291	-0.0016	0.0116	0.5320	0.0008	0.0240	0.5173
Border region	0.1422	0.1835	0.7651	0.0082	0.0237	0.7186	0.1341	0.1797	0.7589
Objective 1 region	0.0377	0.0985	0.6334	0.0038	0.0160	0.6012	0.0339	0.0950	0.6279
Coastal region	-0.0062	0.0500	0.5151	0.0016	0.0131	0.5133	-0.0078	0.0472	0.5240
Coastal region	0.0010	0.0476	0.5356	0.0059	0.0303	0.5546	0.0069	0.0682	0.5450

(b) - Freight attraction

Variable	Average direct impacts			Average indirect impacts			Average total impacts		
	Mean	Std. dev.	Sign prob.	Mean	Std. dev.	Sign prob.	Mean	Std. dev.	Sign prob.
Regional domestic product per capita	-0.0963	0.2008	0.6927	-0.0682	0.1396	0.6925	-0.0282	0.0667	0.6929
Manufacturing gross value added	1.9322	0.4847	0.9999	1.4390	0.4050	0.9999	0.4932	0.2438	0.9999
Population density	0.1766	0.2800	0.7484	0.1235	0.2109	0.7290	0.0531	0.0805	0.7962
Manufacturing employment	-0.0488	0.7487	0.5130	-0.0379	0.5330	0.5135	-0.0109	0.2394	0.5115
Construction employment	0.0269	0.4086	0.5214	0.0216	0.3068	0.5212	0.0052	0.1121	0.5204
Market services employment	0.0673	0.3019	0.5693	0.0484	0.2159	0.5688	0.0189	0.0977	0.5704
Avg. production of manufacturing units	-0.3944	0.2736	0.9101	-0.2948	0.2034	0.9101	-0.0995	0.0886	0.9103
Degree of urbanization	-0.0041	0.0331	0.5518	-0.0031	0.0242	0.5516	-0.0010	0.0098	0.5517
Share of logistic companies	0.6604	1.3842	0.6697	0.5092	1.0620	0.6709	0.1512	0.3744	0.6680
Logistic employment (share)	0.6648	2.1893	0.5986	0.5043	1.6731	0.6009	0.1605	0.6025	0.5926
Number of airports	0.0063	0.0187	0.6267	0.0046	0.0134	0.6269	0.0017	0.0058	0.6247
Length of road network (km)	0.2348	0.4251	0.7361	0.1710	0.3122	0.7364	0.0638	0.1270	0.7357
Number of seaports	-0.0018	0.0088	0.5651	-0.0013	0.0060	0.5658	-0.0006	0.0030	0.5637
Travel time to seaport (min)	0.0459	0.0928	0.6798	0.0342	0.0680	0.6800	0.0117	0.0287	0.6795
Capital city region	-0.4832	0.1341	1.0000	-0.3488	0.0765	1.0000	-0.1344	0.0850	0.9948
Region with large city	0.0309	0.0997	0.6079	0.0079	0.0303	0.5988	0.0230	0.0887	0.6008
Rural region	0.0001	0.0350	0.5158	-0.0019	0.0130	0.5300	0.0020	0.0308	0.5077
Border region	0.0852	0.1495	0.7051	0.0079	0.0242	0.6689	0.0772	0.1452	0.6991
Objective 1 region	0.0514	0.1159	0.6471	0.0038	0.0148	0.6121	0.0477	0.1122	0.6443
Coastal region	-0.0104	0.0534	0.5407	-0.0002	0.0097	0.5188	-0.0102	0.0511	0.5442

Notes: Posterior results based on **W** matrix with 7 nearest neighbours.
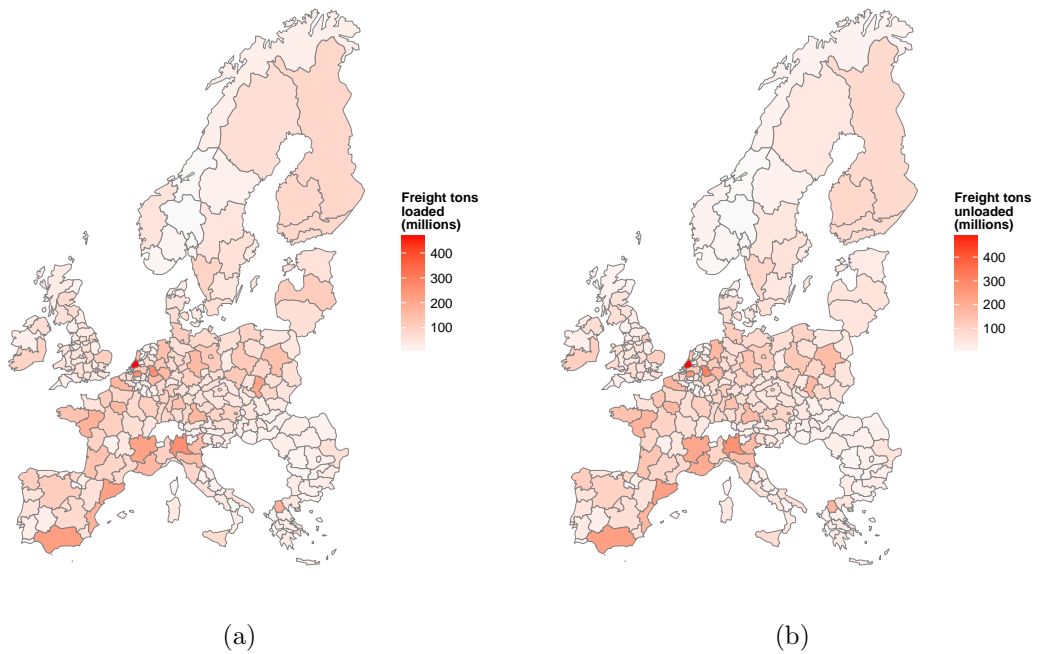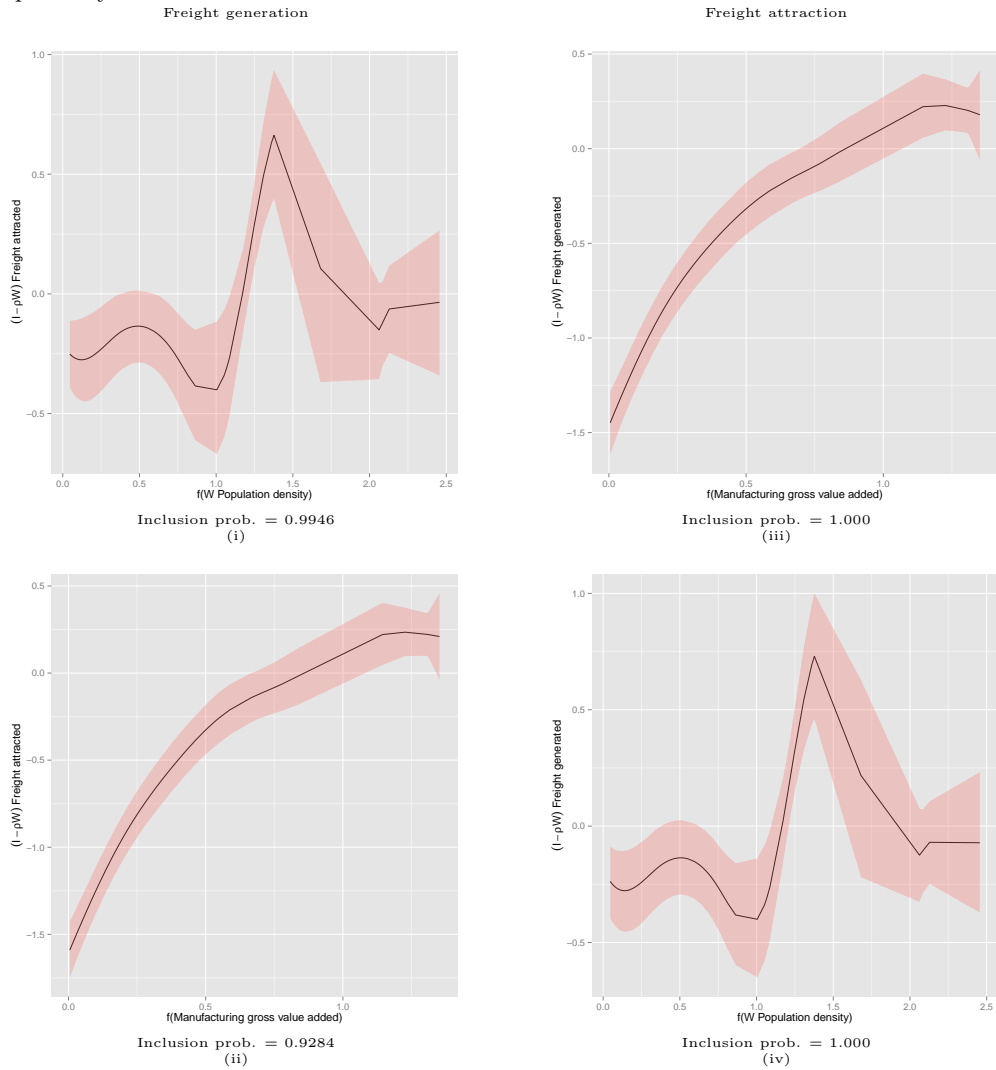
(a)   (b)

Fig. 1: Average yearly million tons of freight generated by (a) and attracted to (b) NUTS-2 regions, 2010 - 2014.

Fig. 2: Functional fit of the penalized spline functions for the two model terms with the highest posterior inclusion probability, in the freight generation (i – ii) and attraction models (iii – iv), respectively.

Appendix A

We assume that the unknown function $f_m(\boldsymbol{v}_m)$ can be approximated by a polynomial splines of degree D_m. $f_m(\cdot)$ is defined over a set of $\boldsymbol{\chi}_m$ knots. The set of knots consist of L_m equally spaced over the range spanned by $min(\boldsymbol{v}_m)$ and $max(\boldsymbol{v}_m)$, and $2D_m$ support knots.

More formally, let $\chi_{m,p_m} \in \boldsymbol{\chi}_m$ be a typical knot, with $p_m = 1, \ldots, P_m$ and $P_m = L_m + 2D_m$. Then $\boldsymbol{\chi}_m$ is constructed so that:

$$\boldsymbol{\chi}_m = [\chi_{m,1}, \ldots, \chi_{m,P_m}] \tag{28}$$
$$\text{where } min(\boldsymbol{v}_m) = \chi_{m,1} = \cdots = \chi_{m,D_m}$$
$$\chi_{m,D_m+1} < \cdots < \chi_{m,L_m+D_m}$$
$$\chi_{m,L_m+D_m+1} = \cdots = \chi_{m,P_m} = max(\boldsymbol{v}_m)-$$

Thus we can approximate the nonlinear function $f_m(\cdot)$ through $L_m = P_m + 2D_m$ basis functions (see the seminal work by DeBoor 1978):

$$f_m(\boldsymbol{v}_m) \approx \bar{\mathbf{Z}}_m \bar{\boldsymbol{\beta}}_m \tag{29}$$

where $\bar{\mathbf{Z}}_m$ is a $n \times P_m$ matrix, where each column corresponds to a basis function and $\bar{\boldsymbol{\beta}}_m$ is the corresponding $P_m \times 1$ vector of parameters.

Now let $\bar{\boldsymbol{z}}_i$ be the i-th row of $\bar{\mathbf{Z}}_m$ and $v_{i,m}$ the i-th element of $\boldsymbol{v}_{i,m}$. Then we can write:

$$f_m(v_{i,m}) \approx \sum_{p_m=1}^{P_m} \bar{\beta}_{m,p_m} B^{D_m}(v_{i,m}) \tag{30}$$

where $B^{D_m}(\cdot)$ denotes a basis function of degree D_m. These basis functions are defined recursively over $d_m = 0, \ldots, D_m$ as (see DeBoor 1978 for further details):

for $d_m = 1, \ldots, D_m$:
$$B^{d_m}(v_{i,m}) = \frac{v_{i,m} - \chi_{m,p_m}}{\chi_{m,p_m+d_m} - \chi_{m,p_m}} B^{d_m-1}(v_{i,m}) + \frac{\chi_{m,p_m+d_m+1} - v_{i,m}}{\chi_{m,p_m+d_m+1} - \chi_{m,p_m+1}} B^{d_m-1}(v_{i,m}) \tag{31}$$
for $d_m = 0$:
$$B^0(v_{i,m}) = \mathbb{1}_{[\chi_{m,p_m}, \chi_{m,p_m+1}]}(v_{i,m}) = \begin{cases} 1 & \chi_{m,p_m} \leq v_{i,m} < \chi_{m,p_m+1} \\ 0 & \text{otherwise} \end{cases} \tag{32}$$

where basis functions of the first degree are denoted by $\mathbb{1}$. These take on values of one between knot points χ_{m,p_m} and χ_{m,p_m+1} otherwise they are zero. Higher order B-splines are defined recursively. Eq. (31) implies that each basis function of degree D_m is defined only over a local interval, i.e. only over the neighbouring $2D_m$ knots.

The derivative of a B-spline is a B-spline itself, comprised of piecewise polynomials. The derivative of a basis function is defined as:

$$\frac{\partial}{\partial v_{i,m}} B^{D_m}(v_{i,m}) = D_m \left(\frac{B^{D_m-1}(v_{i,m})}{\chi_{m,p_m+D_m} - \chi_{m,p_m}} - \frac{B^{D_m-1}(v_{i,m})}{\chi_{m,p_m+D_m+1} - \chi_{m,p_m+1}} \right) \tag{33}$$

which can be expressed as a spline of degree $D_m - 1$.

Appendix B

Table B1: List of countries and NUTS-2 regions

Code	Region name	Code	Region name	Code	Region name	Code	Region name
Austria		**Denmark**		**Italy**		**Sweden**	
AT11	Burgenland (AT)	DK01	Hovedstaden	ITC1	Piemonte	SE11	Stockholm
AT12	Niederösterreich	DK02	Sjælland	ITC2	Valle d'Aosta	SE12	Östra Mellansverige
AT13	Wien	DK03	Syddanmark	ITC3	Liguria	SE21	Småland med öarna
AT21	Kärnten	DK04	Midtjylland	ITC4	Lombardia	SE22	Sydsverige
AT22	Steiermark	DK05	Nordjylland	ITF1	Abruzzo	SE23	Västsverige
AT31	Oberösterreich			ITF2	Molise	SE31	Norra Mellansverige
AT32	Salzburg	**Estonia**		ITF3	Campania	SE32	Mellersta Norrland
AT33	Tirol	EE00	Eesti	ITF4	Puglia	SE33	Övre Norrland
AT34	Vorarlberg	**Greece**		ITF5	Basilicata	**Slovenia**	
Belgium		EL11	Anatoliki Makedonia, Thraki	ITF6	Calabria	SI01	Vzhodna Slovenija
BE10	Région de Bruxelles-Capitale	EL12	Kentriki Makedonia	ITG1	Sicilia	SI02	Zahodna Slovenija
BE21	Prov. Antwerpen	EL13	Dytiki Makedonia	ITG2	Sardegna	**Slovakia**	
BE22	Prov. Limburg (BE)	EL14	Thessalia	ITH1	Provincia Autonoma di Bolzano	SK01	Bratislavský kraj
BE23	Prov. Oost-Vlaanderen	EL21	Ipeiros	ITH2	Provincia Autonoma di Trento	SK02	Západné Slovensko
BE24	Prov. Vlaams-Brabant	EL22	Ionia Nisia	ITH3	Veneto	SK03	Stredné Slovensko
BE25	Prov. West-Vlaanderen	EL23	Dytiki Ellada	ITH4	Friuli-Venezia Giulia	SK04	Východné Slovensko
BE31	Prov. Brabant Wallon	EL24	Sterea Ellada	ITH5	Emilia-Romagna	**United Kingdom**	
BE32	Prov. Hainaut	EL25	Peloponnisos	ITI1	Toscana	UKC1	Tees Valley, Durham
BE33	Prov. Liège	EL30	Attiki	ITI2	Umbria	UKC2	Northumberland, Tyne and Wear
BE34	Prov. Luxembourg (BE)	EL41	Voreio Aigaio	ITI3	Marche	UKD1	Cumbria
BE35	Prov. Namur	EL42	Notio Aigaio	ITI4	Lazio	UKD3	Greater Manchester
Bulgaria		EL43	Kriti	**Latvia**		UKD4	Lancashire
BG31	Severozapaden	**Spain**		LT00	Lietuva	UKD6	Cheshire
BG32	Severen tsentralen	ES11	Galicia	**Luxembourg**		UKD7	Merseyside
BG33	Severoiztochen	ES12	Principado de Asturias	LU00	Luxembourg	UKE1	East Yorkshire and Northern Lincolnshire
BG34	Yugoiztochen	ES13	Cantabria	**Lithuania**		UKE2	North Yorkshire
BG41	Yugozapaden	ES21	País Vasco	LV00	Latvija	UKE3	South Yorkshire
BG42	Yuzhen tsentralen	ES22	Comunidad Foral de Navarra	**Netherlands**		UKE4	West Yorkshire
Czech Republic		ES23	La Rioja	NL11	Groningen	UKF1	Derbyshire, Nottinghamshire
CZ01	Praha	ES24	Aragón	NL12	Friesland (NL)	UKF2	Leicestershire, Rutland and Northamptonshire
CZ02	Strední Cechy	ES30	Comunidad de Madrid	NL13	Drenthe	UKF3	Lincolnshire
CZ03	Jihozápad	ES41	Castilla y León	NL21	Overijssel	UKG1	Herefordshire, Worcestershire and Warwickshire
CZ04	Severozápad	ES42	Castilla-la Mancha	NL22	Gelderland	UKG2	Shropshire, Staffordshire
CZ05	Severovýchod	ES43	Extremadura	NL23	Flevoland	UKG3	West Midlands
CZ06	Jihovýchod	ES51	Cataluña	NL31	Utrecht	UKH1	East Anglia
CZ07	Strední Morava	ES52	Comunidad Valenciana	NL32	Noord-Holland	UKH2	Bedfordshire, Hertfordshire
CZ08	Moravskoslezsko	ES53	Illes Balears	NL33	Zuid-Holland	UKH3	Essex
Germany		ES61	Andalucía	NL34	Zeeland	UKI1	Inner London
DE11	Stuttgart	**Finland**		NL41	Noord-Brabant	UKI2	Outer London
DE12	Karlsruhe	FI19	Länsi-Suomi	NL42	Limburg (NL)	UKJ1	Berkshire, Buckinghamshire and Oxfordshire
DE13	Freiburg	FI1B	Helsinki-Uusimaa	**Poland**		UKJ2	Surrey, East, West Sussex
DE14	Tübingen	FI1C	Etelä-Suomi	PL11	Lódzkie	UKJ3	Hampshire, Isle of Wight
DE21	Oberbayern	FI1D	Pohjois- ja Itä-Suomi	PL12	Mazowieckie	UKJ4	Kent
DE22	Niederbayern	**France**		PL21	Malopolskie	UKK1	Gloucestershire, Wiltshire, Bristol
DE23	Oberpfalz	FR10	Île de France	PL22	Slaskie	UKK2	Dorset, Somerset
DE24	Oberfranken	FR21	Champagne-Ardenne	PL31	Lubelskie	UKK3	Cornwall, Isles of Scilly
DE25	Mittelfranken	FR22	Picardie	PL32	Podkarpackie	UKK4	Devon
DE26	Unterfranken	FR23	Haute-Normandie	PL33	Swietokrzyskie	UKL1	West Wales, The Valleys
DE27	Schwaben	FR24	Centre (FR)	PL34	Podlaskie	UKL2	East Wales
DE30	Berlin	FR25	Basse-Normandie	PL41	Wielkopolskie	UKM2	Eastern Scotland
DE40	Brandenburg	FR26	Bourgogne	PL42	Zachodniopomorskie	UKM3	South Western Scotland
DE50	Bremen	FR30	Nord - Pas-de-Calais	PL43	Lubuskie	UKM5	North Eastern Scotland
DE60	Hamburg	FR41	Lorraine	PL51	Dolnoslaskie	UKM6	Highlands, Islands
DE71	Darmstadt	FR42	Alsace	PL52	Opolskie	UKN0	Northern Ireland (UK)
DE72	Gießen	FR43	Franche-Comté	PL61	Kujawsko-Pomorskie		
DE73	Kassel	FR51	Pays de la Loire	PL62	Warminsko-Mazurskie		
DE80	Mecklenburg-Vorpommern	FR52	Bretagne	PL63	Pomorskie		
DE91	Braunschweig	FR53	Poitou-Charentes	**Portugal**			
DE92	Hannover	FR61	Aquitaine	PT11	Norte		
DE93	Lüneburg	FR62	Midi-Pyrénées	PT15	Algarve		
DE94	Weser-Ems	FR63	Limousin	PT16	Centro (PT)		
DEA1	Düsseldorf	FR71	Rhône-Alpes	PT17	Área Metropolitana de Lisboa		
DEA2	Köln	FR72	Auvergne	PT18	Alentejo		
DEA3	Münster	FR81	Languedoc-Roussillon	**Romania**			
DEA4	Detmold	FR82	Provence-Alpes-Côte d'Azur	RO11	Nord-Vest		
DEA5	Arnsberg	FR83	Corse	RO12	Centru		
DEB1	Koblenz	**Hungary**		RO21	Nord-Est		
DEB2	Trier	HU10	Közép-Magyarország	RO22	Sud-Est		
DEB3	Rheinhessen-Pfalz	HU21	Közép-Dunántúl	RO31	Sud - Muntenia		
DEC0	Saarland	HU22	Nyugat-Dunántúl	RO32	Bucuresti - Ilfov		
DED2	Dresden	HU23	Dél-Dunántúl	RO41	Sud-Vest Oltenia		
DED4	Chemnitz	HU31	Észak-Magyarország	RO42	Vest		
DED5	Leipzig	HU32	Észak-Alföld				
DEE0	Sachsen-Anhalt	HU33	Dél-Alföld				
DEF0	Schleswig-Holstein	**Ireland**					
DEG0	Thüringen	IE01	Border, Midland, Western				
		IE02	Southern, Eastern				

Table B2: Summary statistics for the covariates

Variable	Mean	Std. dev.	Min	Max
Freight generated (1,000t)	0.66	0.55	0.01	4.66
Freight attracted (1,000t)	0.69	0.58	0.02	4.99
Regional domestic product per capita	0.22	0.26	0.01	2.56
Manufacturing gross value added	0.21	0.22	0.01	1.35
Population density	0.34	0.85	0.01	9.07
Manufacturing employment	0.08	0.02	0.03	0.16
Construction employment	0.26	0.04	0.13	0.48
Market services employment	0.31	0.06	0.12	0.45
Avg. production of manufacturing units	0.48	0.61	0.00	3.12
Degree of urbanization	0.40	1.11	0.00	17.28
Share of logistic companies	0.05	0.03	0.00	0.14
Logistic employment (share)	0.03	0.01	0.00	0.12
Number of airports	1.59	1.78	0.00	15.00
Length of road network (km)	0.15	0.15	0.00	1.05
Number of seaports	2.25	4.54	0.00	34.00
Travel time to seaport (min)	0.59	0.83	0.01	3.00
Capital city region	0.16	0.37	0.00	1.00
Region with large city	0.72	0.45	0.00	1.00
Rural region	0.66	0.48	0.00	1.00
Border region	0.53	0.50	0.00	1.00
Objective 1 region	0.34	0.47	0.00	1.00
Coastal region	0.47	0.50	0.00	1.00

Table B3: Non-semi-parametric freight generation (a) and attraction (b) model spatial impact estimates

(a) - Freight generation

Variable	Average direct impacts			Average indirect impacts			Average total impacts		
	Mean	Std. dev.	Sign prob.	Mean	Std. dev.	Sign prob.	Mean	Std. dev.	Sign prob.
Regional domestic product per capita	0.5227	0.4611	0.8820	-0.2584	0.1263	0.9800	0.7810	0.4317	0.9660
Manufacturing gross value added	1.0620	0.6728	0.9470	1.4826	0.1667	1.0000	-0.4206	0.6073	0.7550
Population density	0.2853	0.2230	0.8650	-0.0362	0.0426	0.8080	0.2895	0.1945	0.9280
Manufacturing employment	-6.1242	3.8015	0.9460	1.1093	1.8922	0.7190	-7.2336	3.9169	0.9720
Construction employment	1.3936	2.5506	0.7190	-0.4469	0.8648	0.6990	1.8405	2.4382	0.7710
Market services employment	0.5137	1.5770	0.6290	0.8850	0.5931	0.9430	-0.3714	1.5778	0.5830
Avg. production of manufacturing units	-0.1883	0.1753	0.8650	0.1389	0.0752	0.9700	-0.3272	0.1738	0.9700
Degree of urbanization	0.1445	0.1255	0.8640	0.0195	0.0244	0.7890	0.1250	0.1168	0.8510
Share of logistic companies	2.5412	3.2545	0.7740	-0.2835	1.1664	0.5960	2.8247	3.2038	0.8190
Logistic employment (share)	-12.3022	9.1610	0.9110	1.3584	2.1920	0.7350	-13.6606	8.5535	0.9410
Number of airports	0.0427	0.0689	0.7300	0.0329	0.0191	0.9580	0.0099	0.0653	0.5670
Length of road network (km)	0.4796	0.7977	0.7350	0.2655	0.1968	0.9080	0.2141	0.7569	0.6210
Number of seaports	-0.0063	0.0290	0.5770	0.0032	0.0065	0.6840	-0.0095	0.0265	0.6320
Travel time to seaport (min)	-0.0947	0.1532	0.7320	-0.0389	0.0396	0.8330	-0.0559	0.1513	0.6430
Capital city region	-0.5632	0.3144	0.9650	-0.3689	0.0901	1.0000	-0.1943	0.2912	0.7510
Region with large city	0.5997	0.3080	0.9760	0.1669	0.0688	0.9890	0.4328	0.2840	0.9380
Rural region	-0.0067	0.2025	0.5160	-0.1173	0.0683	0.9570	0.1106	0.1901	0.7160
Border region	0.3782	0.1661	0.9920	0.0177	0.0693	0.6060	0.3605	0.1786	0.9810
Objective 1 region	0.3562	0.1807	0.9800	0.0499	0.0717	0.7580	0.3063	0.1853	0.9500
Coastal region	0.1115	0.2756	0.6360	0.0544	0.0747	0.7580	0.0571	0.2675	0.5650
σ^2							0.1441	0.0133	
ρ							0.2819	0.0117	

(b) - Freight attraction

Variable	Average direct impacts			Average indirect impacts			Average total impacts		
	Mean	Std. dev.	Sign prob.	Mean	Std. dev.	Sign prob.	Mean	Std. dev.	Sign prob.
Regional domestic product per capita	0.5208	0.4473	0.8810	-0.2583	0.1302	0.9810	0.7792	0.4250	0.9680
Manufacturing gross value added	1.2339	0.7891	0.9390	1.5841	0.1875	1.0000	-0.3502	0.7068	0.6810
Population density	0.2853	0.2480	0.8640	-0.0114	0.0462	0.5940	0.2968	0.2172	0.9120
Manufacturing employment	-7.1104	4.2380	0.9560	0.8745	1.9370	0.6820	-7.9849	4.1956	0.9710
Construction employment	1.4565	2.6305	0.7160	-0.4045	0.8904	0.6690	1.8610	2.5620	0.7720
Market services employment	0.2416	1.6251	0.5650	1.0884	0.6100	0.9600	-0.8468	1.6560	0.6940
Avg. production of manufacturing units	-0.2103	0.1820	0.8760	0.1752	0.0785	0.9880	-0.3855	0.1799	0.9890
Degree of urbanization	0.1580	0.1374	0.8810	0.0266	0.0243	0.8680	0.1314	0.1272	0.8500
Share of logistic companies	2.6712	3.4266	0.7800	-0.3174	1.2766	0.6150	2.9887	3.3120	0.8230
Logistic employment (share)	-10.3603	9.5122	0.8650	1.3804	2.3226	0.7200	-11.7407	8.7803	0.9160
Number of airports	0.0343	0.0748	0.6710	0.0275	0.0201	0.9160	0.0068	0.0711	0.5410
Length of road network (km)	0.4457	0.8132	0.7170	0.2500	0.1960	0.8950	0.1957	0.7744	0.5860
Number of seaports	-0.0054	0.0298	0.5690	0.0071	0.0068	0.8490	-0.0125	0.0277	0.6700
Travel time to seaport (min)	-0.1522	0.1634	0.8180	-0.0256	0.0403	0.7440	-0.2410	0.3236	0.7730
Capital city region	-0.6323	0.3524	0.9600	-0.3913	0.0971	0.9990	0.4786	0.2816	0.9530
Region with large city	0.6549	0.3096	0.9820	0.1763	0.0743	0.9880	0.0823	0.1929	0.6680
Rural region	-0.0146	0.2028	0.5280	-0.0969	0.0708	0.9250	0.3875	0.1872	0.9800
Border region	0.3998	0.1784	0.9860	0.0124	0.0728	0.5580	0.3212	0.2029	0.9450
Objective 1 region	0.3546	0.2013	0.9600	0.0334	0.0743	0.6760	0.0059	0.2772	0.5240
Coastal region	0.1139	0.2872	0.6550	0.1080	0.0780	0.9260			
σ^2							0.1538	0.0136	
ρ							0.3508	0.0329	

Notes: Posterior results based on **W** matrix with 7 nearest neighbours.